Source Associations For The Virtual Observatory

Emma Taylor

Institute for Astronomy
School of Physics

University of Edinburgh
Doctor of Philosophy

December 2005

Abstract

This thesis presents investigations into different methods of associating astronomical sources detected at different wavelengths, and describes the development of a tool for AstroGrid to enable users to associate sources in a fully automated manner.

At present when associating sources at different wavelengths it is common for astronomers to select IDs by eye or at least verify probabilistically–determined counterparts by eye. With the new trend for large surveys this is no longer practical as datasets may contain millions of objects. Previous work on association algorithms has focussed on case–specific techniques which typically only match a restricted number of objects with counterparts, and often only those with small positional errors. This thesis addresses the issue that these methods are not adequate in the general case where datasets may be enormous and source error ellipses large. In such situations matching based purely on spatial proximity is deficient since there may be hundreds of candidate counterparts within a source error ellipse. We therefore investigate the likelihood ratio as an association technique, as this allows incorporation of data such as object magnitudes as well as positions, and prove its applicability in the (difficult association) case of the FIRBACK survey. We also develop the application of a machine learning technique, the EM algorithm, and test it against the likelihood ratio method. We determine that it may be effectively applied to find IDs in surveys with a magnitude distribution with unrestricted shape. These different association methods are successfully developed into a tool for AstroGrid to enable users to associate sources in a fully automated manner.

We describe detailed analysis of the likelihood ratio method through the association of a population of far–infrared sources from the FIRBACK survey with optical counterparts from the INT Wide Field Survey. This is a challenging association problem since the far–infrared sources have a large positional error due to the poor resolution of the instrument and the relatively long wavelength. We compare two different variants of the likelihood ratio method in detail, and use the better one to derive optical counterparts for the far–infrared sources. This proves the applicability of the likelihood ratio

method in the case of large source error ellipses where there are numerous candidates to choose between.

The scientific benefits of associating multiwavelength data are illustrated via deducing, for the first time, the nature of the FIRBACK sources. These are identified with not only an optical counterpart but also with data at up to nine further wavelengths. Their properties are examined through the comparison of their observed spectral energy distributions with predictions from radiative transfer models which simulate the emission from both cirrus and starburst components. The far–infrared sources are found to be 80 per cent star–bursting galaxies with their starburst component at a high optical depth.

It is a common situation in astronomy to wish to investigate a source population for which we have no prior knowledge about the properties of the source counterparts expected at another wavelength, for example through observations with a new instrument. In such a case it is necessary to estimate the counterpart magnitude distribution to use the likelihood ratio association method. Since little was known about the FIRBACK sources, prior to our research, their optical magnitude distribution had to be estimated in order to assign them optical IDs. To alleviate this problem we develop a new astronomical application of a machine learning technique known as the EM algorithm which is used in the field of informatics. This is able to 'learn' the source magnitude distribution iteratively. The algorithm is tested on the FIRBACK sources and also radio sources from the HI Parkes All–Sky Survey (HIPASS) catalogue and is found to be a very effective association method in the HIPASS case where the background magnitude distribution is of unrestricted shape.

We use the FIRBACK survey far–infrared sources as a test–bed for several different association methods. The value of bringing together multiwavelength observations is illustrated through the insights that are gained into the nature of the sources. This work culminates in the development of an association tool for AstroGrid, the UK Virtual Observatory project, offering three different association methods: the Poisson method, the likelihood ratio method and the EM algorithm. This tool is able to return a user specified number of possible counterparts along with a figure of merit for their match with a source. We also implement the AstroDAS system to store resulting object pairs in a database for future use. This prevents the same cross association tasks being carried out numerous times by different users. The Virtual Observatory aims to link diverse datasets from across the globe. The extra knowledge available from these may only be extracted after establishing links between detections in these datasets. Our

AstroGrid association tool is therefore vital to the success of the Virtual Observatory.

Declaration

I declare that this thesis is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other University. I further state that no part of my thesis has already been or is being concurrently submitted for any such degree, diploma or other qualification.

This thesis is the outcome of my own work except where specifically indicated in the text.

Emma Taylor
Edinburgh,
December 2005.

Acknowledgements

I would first of all like to thank my supervisor Bob for guiding me throughout this work. Chapter 4 would also not have been possible without the invaluable input of Amos Storkey who taught me all about the EM Algorithm and thank you to David Rohde et al. for the use of their data. I am greatly indebted to John Taylor and Martin Hill for all their help and direction with integrating my tool into AstroGrid and for helping me to learn Java.

Thank you also to all the other students at the ROE for their random chat at coffee times which kept me sane through those 1st, 2nd and 3rd year blues. I would particularly like to mention Rachel Dowsett who was the most understanding office mate anyone could have and Jessica Skelton, Michael Davidson and Miller Crawford. Thank you to Anita Schael and Michele Cirasuolo for introducing me to salsa which although great for stress relief was probably not the greatest thing to become addicted to in the final months of writing up!

The greatest thanks of all must go to my wonderful husband Richard who listened to all my moaning and gave me the motivation to finish so we could get married. He was with me through all the difficult times when I thought I would never finsh writing my paper, my algorithms would never work and I had lost all enthusiasm. Thank you for listening to me on the phone for hours every day. Especially thank you for teaching me about HII emission before my viva! You are my inspiration.

These acknowledgements would not be complete without thanking Cafe Nero on Morningside Road who supplied me with copious amounts of latte and where a large proportion of this thesis was written. And also thanks to my Dad for making sure that this thesis tells a story!

Finally, thank you to God who was with me and supported me throughout my time in Edinburgh. I know I couldn't have got through the viva without Him.

Contents

1 Introduction **1**

1.1 Multiwavelength Astronomy 1

1.2 The Virtual Observatory Concept 3

1.3 Building the Virtual Observatory 4

1.4 The Virtual Observatory Architecture 5

1.5 Current Status of the Virtual Observatory 7

1.6 Associations and the Virtual Observatory 8

1.7 Review of Previous Association Methods 9

1.7.1 The Poisson Model 10

1.7.2 The Likelihood Ratio Method 13

1.8 Plan of the Thesis 14

2 The Likelihood Ratio Association Algorithm **15**

2.1 Introduction 15

2.2 The Far-Infrared Background 16

2.3 Current Thoughts on the Nature of the FIRBACK Sources 17

2.4 Current Competing Evolutionary Models 18

2.5 The Data 20

2.5.1 The FIRBACK Survey 20

2.5.2 The INT Wide Field Survey (WFS) 20

2.6 Likelihood Ratio Association Algorithms 20

2.6.1 The Sutherland and Saunders (1992) Method 20

2.6.2 The Rutledge et al. (2000) Method 22

2.6.3 The Mann et al. (1997, 2002) Method 25

2.7 Implementation of Association Algorithms 26

2.7.1 The Rutledge et al. (2000) Method 26

2.7.2 The Mann et al. (1997, 2002) Method 35
2.8 Summary and Discussion 41
2.9 Associating the N2 FIRBACK Optical IDs with further multiwavelength data 43

3 The Nature of the 175μm Selected N2 FIRBACK Sources 47
3.1 Introduction 47
3.2 Optical Properties 48
3.3 Radiative Transfer Models 50
3.4 Spectral Energy Distributions 54
3.5 Star Formation Rates and Far-Infrared Luminosities 56
3.6 Dust Temperatures and Masses 59
3.7 The Nature of the N2 FIRBACK Sources 68
3.8 Summary of the Nature of the N1 FIRBACK Sources 70

4 Machine Learning and Associations 73
4.1 Introduction 73
4.2 FIRBACK Associations with the EM Algorithm 74
4.2.1 Probability Distribution in the Data 75
4.2.2 Setting up the EM Algorithm 76
4.3 Associations for the N1 FIRBACK Sources Using the Likelihood Ratio and EM Algorithm 80
4.4 Using Machine Learning Methods to Obtain HIPASS Associations . . . 88
4.4.1 The HIPASS Survey 88
4.4.2 The SuperCOSMOS Sky Survey 89
4.4.3 Support Vector Machines and Neural Networks 89
4.4.4 Support Vector Machines and Neural Networks as a Matching Tool (Rohde et al. 2005) 92
4.4.5 Associations for the HIPASS sources using the Likelihood Ratio Method and the EM Algorithm 94
4.5 Comparison of Associations Made for the HIPASS Sources 99
4.6 Comparison of the Performance of the EM Algorithm on the Two Datasets 109
4.7 Summary 111
4.8 Conclusions 117

5 Implementing Association Algorithms On AstroGrid **119**

5.1 Introduction 119

5.2 AstroGrid Architecture 120

5.3 Review of other VO Association Work 125

5.4 Development of an Association Tool for AstroGrid 131

5.5 Portal Tour and Application Testing 134

5.6 Using the xmatch Tool on AstroGrid 143

5.7 AstroDAS 148

5.8 Conclusions 155

6 Conclusions **159**

6.1 The Need for an Association Tool in the Virtual Observatory 159

6.2 The Likelihood Ratio 160

6.3 The Nature of the 175 μm FIRBACK Sources 162

6.4 The EM Algorithm 163

6.5 An Association Tool for the VO 164

6.6 Conclusions and Future Work 166

A FIRBACK N2 IDs **175**

B Published Papers **195**

B.1 Published Papers 195

My son, there is something else to watch out for. There is no end to the writing of books, and too much study will wear you out.

Ecclesiastes 12 v.12

1 Introduction

1.1 Multiwavelength Astronomy

Astronomers observe the sky at wavelengths covering the full electromagnetic spectrum (16 orders of magnitude in wavelength) from gamma rays to radio waves. Methods for detecting photons of such widely varying energies are very different: For example at optical wavelengths images are formed through the accumulation of photons into CCD pixels; X-ray photons are so rare that individual photons are recorded and radio wavelength photons are detected via interferometers producing a Fourier–transform of the sky. The data obtained at these different wavelengths are therefore also very different. This has led to astronomers traditionally being experts at a particular wavelength since they were familiar with one type of instrument, data and analysis method. Therefore separate data centres emerged with expertise in different regions of the EM spectrum (Mann and Williams 2004)[1].

However, over the past 20–30 years the benefits of multiwavelength astronomy have become apparent and astronomers have begun to become experts in particular types of celestial object rather than all objects in one region of the spectrum. The physical

[1] 'What is the Virtual Observatory?': http://www.roe.ac.uk/~rgm/sc4devo/WhatIsTheVO.pdf

mechanisms that produce photons of such widely varying energies are very different as is the spatial distribution of those photons, as illustrated in Figure 1.1, therefore to understand the properties of an astronomical source requires the construction of a model which can explain all its emission (Mann and Williams 2004). This therefore requires the astronomer to bring together observations at many wavelengths and provides the motivation for federated astronomy. Illustrations of this are the search for brown dwarfs which require both optical and infrared data (e.g. Berriman et al. 2003) and the study of quasars which requires data at all wavelengths (e.g. Padovani et al. 2004).

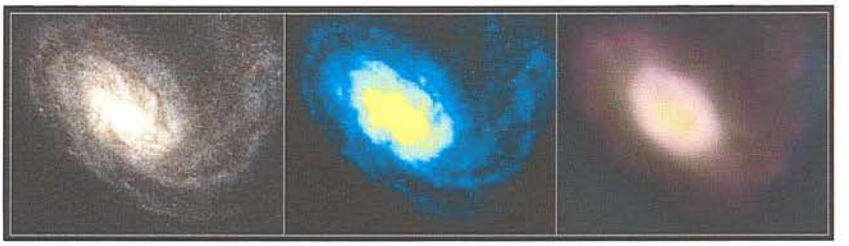

Figure 1.1: Images of the same galaxy taken at different wavelengths: optical (left), radio (middle) and X-ray (right). (Images taken from http://euro-vo.org/)

This shift to federated astronomy has meant that obtaining data has had to become simpler if an astronomer is to use a wide variety of data. Therefore common user instruments have been introduced at observatories to make it easier for an astronomer observing in a new band. Usually these also provide a data reduction pipeline so that reduction is quick and simple and the astronomer can start the science as soon after observation as possible. This has led to science–ready data products being made available through archives and as astronomers have begun to realise the benefits of the use of such data this has in turn led to an enormous expansion in survey astronomy (Mann and Williams 2004). Surveys produce vast amounts of homogeneous and well characterised data stored in organised databases. Examples of these include object catalogues containing $10^8 - 10^9$ object entries such as the all–sky photographic optical USNO–B1 catalogue (Monet et al. 2003), the all–sky near–IR 2MASS catalogue (Jarrett et al. 2000), the wide field Sloan Digital Sky Survey (SDSS) (Abazajian et al. 2005) and the SuperCOSMOS Science Archive (SSA) (Hambly et al. 2004) which are each approximately 1TB in size. There are also several galaxy redshift catalogues containing $10^5 - 10^6$ entries such as the 2-degree field (2dF) (Colless et al. 2001) and SDSS Spectroscopic surveys (Abazajian et al. 2005) and source catalogues of similar size for

example the NRAO VLA Sky Survey (Condon et al. 1998) and the IRAS faint source catalogue (Moshir et al. 1990) to name but a few. Many surveys now make their data available to astronomers on–line for example NED (http://ned.ipac.caltech.edu) and SIMBAD (http://simbad.u-strasbg.fr) (Feigelson and Babu 2003).These provide some simple tools but are mostly intended for file download. However, this is becoming increasingly impractical as the amount of data doubles every year and we are set for a data explosion as new instruments come on–line. The recently installed WFCAM can produce 100 GB/night (30 TB/year), VISTA will produce 500 GB/night (200 TB/year) and e–MERLIN/ALMA 1 TB/day (Gonzalez–Solares 2005)[2]. These surveys at different wavelengths are obviously key to gaining insight about an astronomical object through combining the multiwavelength data. However, at present this is not trivial since these databases are heterogeneous in character, located at widely dispersed sites and are accessed through different systems.

1.2 The Virtual Observatory Concept

A substantial fraction of astronomical research papers is now based on archive data. These data may be used in support of other observations, for direct research through off–line analysis of specific data sets or for discovery–based programmes. A 'power–user' implements a discovery programme running statistical analysis of huge amounts of data, re–reducing data in novel ways or searching for rare objects, etc. Such work is at present the domain of specialists who have privileged access to datasets, detailed technical knowledge and put in months of work. This is a new style of science requiring the searching, filtering, manipulation or browsing of entire vast datasets which is a 'data–mining' style of working. The UK AstroGrid Virtual Observatory project, begun in September 2001 and still in development, aims to develop hardware, architecture and data management systems around a number of expert data centres, along with query, analysis and exploration tools enabling everyone to be a power–user from their own desktop and is indeed well on its way towards that goal (AstroGrid Proposal 2001)[3].

The new generation of databases produced by new survey instruments will motivate more and more astronomers to become power users. New avenues of science will become available to them via trawling the multidimensional data spaces made available by database federation. AstroGrid will make this easier, faster and more standard allowing

[2]Presentation at MSSL 24/5/05: http://wiki.astrogrid.org/bin/view/Astrogrid/MeetingsPresentations2005

[3]AstroGrid Proposal to PPARC 2/4/01: http://wiki.astrogrid.org/pub/Astrogrid/CoreDocs/astrogrid-proposal.doc

the astronomer to browse and query many databases from a single entry point. They will be able to log-on to AstroGrid and issue joint queries automatically without having to know where the different databases are held or having to learn different packages and this will all be fast enough to be interactive (AstroGrid Proposal 2001).

1.3 Building the Virtual Observatory

The AstroGrid project is one of many 'Virtual Observatory' (VO) projects worldwide which will provide a set of co–operating and interoperable software systems that:

- allow users to interrogate multiple data centres in a seamless and transparent way
- provide powerful new analysis and visualisation tools
- give data centres a standard framework for publishing services and using their data (AstroGrid Phase A Report 2002)[4].

This will be achieved through the standardisation of data, metadata, data exchange methods and the use of a registry i.e. it will 'make all archives speak the same language' and they will then be searchable and analysable by the same tools. The VO will allow data sources to be accessible through a uniform interface and although they are held at distributed locations they will appear as one. This will form the Digital Sky. Once these systems are in place the VO will make doing astronomy faster and more effective since the astronomer will be freed from many mundane tasks. Interdisciplinary barriers will be removed and distributed collaborative teams formed while resources are pooled and no longer dependent on an astronomer's location, encouraging data–intensive exploratory science (AstroGrid Phase A Report 2002).

The VO can only be successful if it is global and all the world's significant data resources are federated into the VO. National and regional VO projects such as the UK AstroGrid project are all coordinated by the International VO Alliance (IVOA). This currently has 14 members, shown in Figure 1.2, which have all signed up to the IVOA roadmap for the first phase of global VO development (2002-2005) accepting the standards set down by the IVOA and its mission statement: 'To facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organisational structures necessary to enable the international utilisation of astronomical archives as an integrated interoperating Virtual Observatory'[5].

[4]AstroGrid Phase A Report: http://wiki.astrogrid.org/pub/Astrogrid/PhaseAReport/redbook.pdf
[5]IVOA mission statement: http://www.ivoa.net/pub/info/Statement2002-2005.htm

This will ensure that the underlying infrastructure, protocols and software elements are universal (Mann and Williams 2004).

Figure 1.2: Members of the IVOA. (Taken from http://www.ivoa.net/pub/members/)

1.4 The Virtual Observatory Architecture

At a meeting of the IVOA in January 2003 six major technical areas were identified to develop the Virtual Observatory's scientific goals. These were:

Registries are the 'yellow pages' of the VO, listing all the available resources and services. Registries all over the world 'harvest' each other, so that each one holds descriptions of all the resources across the globe.

Data Models define astronomical data structures and their relationships. Although there are common formats in astronomy, such as FITS (Hanisch et al. 2001), there are currently many variations in the metadata.

Uniform Content Descriptors (UCDs) provide a common vocabulary for metadata definitions. This means that it is possible to submit a query for 'PHOT_JHN_V' and have returned the Johnson V magnitude no matter which database is being

queried. UCDs were developed by CDS in Strasbourg to describe the data in their large collection of catalogues and tables. They found that among the tens of thousands of column names being used to describe their data there were only about 1500 unique terms, for example there were 250 different labels in use for Johnson V magnitude.

Data Access Layer adds interfaces and protocols to the Data Models and UCDs to provide a standard method of access to distributed data.

VO Query Language is derived from SQL and provides a standard Astronomical Data Query Language (ADQL) to query any and all of the databases distributed throughout the VO.

VOTable is an XML mark–up standard for astronomical tables and was the first international agreement reached by the collaborating global VO projects. It provides a standard format for passing data around in the VO (Hanisch and Quinn)[6].

These areas of research and development have contributed to the architecture design of the Virtual Observatory. A schematic representation of the UK's AstroGrid architecture is given in Figure 1.3. This is most easily explained if we follow the process initiated by an AstroGrid user, marked by the numbers in the Figure. Firstly the astronomer will login to AstroGrid, using a username, password and community, to one of the three available client applications: a command line interface (CLI), a web Portal or the Workbench.

1. First the user constructs a workflow which sets out the required tasks. This may involve a series of steps done one after the other, in parallel, loops or even conditions. The workflow is then submitted to the system to be executed remotely. Execution is managed by the Job Execution Service (JES) which monitors the status of the job and constantly returns up–to–date information regarding its progress.

2. The JES speaks to the Registry which tells it where to find the services that provide the information requested.

3. Once the JES knows where to look it sends a request to the web service interface of the relevant data set or tool.

[6]'The International Virtual Observatory': http://www.ivoa.net/pub/info/#documents

4. Users are members of a community, and this enables access rights to certain datasets to be restricted to authorised people. When a request is received by a web service interface it will check that the user is allowed access by communicating with the community service.

5. Once the user has been authenticated and authorised then the requested information is assimilated or tool executed and the results are sent to MySpace. This is a storage service where the astronomer can store queries, workflows and results.

This is just an example of how one might use AstroGrid. However, each of the components can be used individually without the use of the other components. This versatility enables astronomers to extract information in the way they would like to.

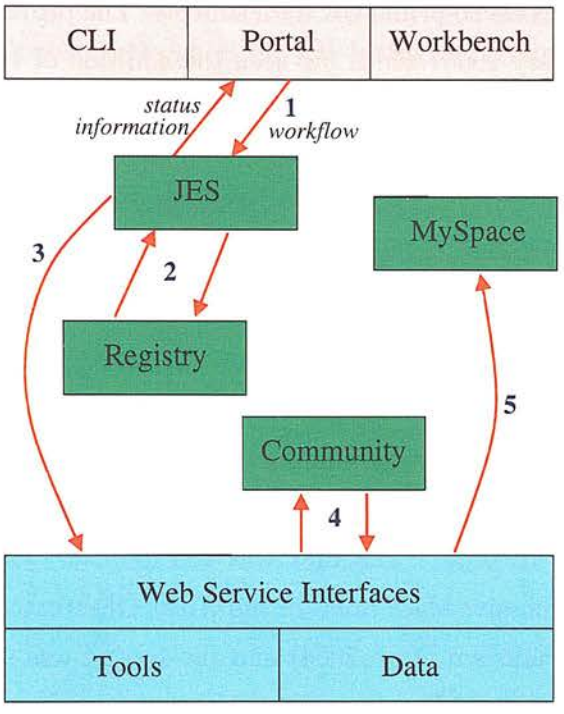

Figure 1.3: Schematic representation of AstroGrid architecture.

1.5 Status of the Virtual Observatory

European astronomers have formed a collaboration of 10 organisations to form a European VO project, Euro–VO, to build an implementation of the VO that will benefit astronomers Europe wide regardless of location. Within this umbrella project are 3 interlinked projects: 1. Euro–VO Data Centre Alliance (DCA) - a network of data centres

publishing data, metadata and services to the Euro–VO 2. Euro–VO Facility Centre (VOFC) - a centralised registry for standards, resources and certification 3. Euro–VO Technology Centre (VOTC) - to coordinate research and development projects for the advancement of VO technology, systems and tools which has implemented a VO–Tech project to develop the VOTC. Initial implementation of Euro–VO is underway in 2005 (VO–Tech Proposal 2004)[7].

The UK AstroGrid project began in September 2001 with an initial phase (Phase A) focusing on analysis, research and development and the architecture design. The initial partners were Edinburgh, Cambridge, Leicester, Rutherford Appleton Lab (RAL), Mullard Space Science Lab (MSSL), Jodrell Bank Observatory and Queens, Belfast. Phase B was started in January 2003 and aimed to build, test and deliver a working infrastructure giving access to prime UK data sources. The project is currently in stage 2 (AG2 began in January 2005) which has seen the addition of 4 new partners - Leeds, Bristol, Exeter and Portsmouth. This stage has aims to improve VO infrastructure, especially user tools, research and develop new techniques and encourage VO uptake to increase the DCA. AstroGrid released its first working system on 1st May 2005 which is publicly available. This release has access to a small set of UK databases and 6000 catalogues, it also provides two packaged science systems - a Solar Movie Maker and Redshift Maker (AstroGrid-2 Proposal 2003)[8].

Over 400 papers have already been published relating to Virtual Observatories proving their potential to enable new astronomical research (IVOA Status Report 2005)[9]. There are two examples of major scientific discoveries that have been facilitated by Virtual Observatories to date. The first was the discovery of 31 previously undetected powerful supermassive black holes in the Great Observatories Origins Deep Survey (GOODS) field (Padovani et al. 2004) and the second was the classification of all unidentified ROSAT WGACAT objects using VO data (McGlynn et al. 2004).

1.6 Associations and the Virtual Observatory

The success of the VO depends heavily on the ease with which science can be extracted from the available data. AG2 plans to provide fully–functioning data exploration ca-

[7]VO–Tech proposal to EU 03/04: http://wiki.astrogrid.org/pub/AG2/WebHome/VO-TECH-PARTB.pdf

[8]AstroGrid-2 Proposal to PPARC 04/03: http://wiki.astrogrid.org/pub/AG2/WebHome/AG2-double.pdf

[9]IVOA Status Report 2005: http://www.ivoa.net/pub/info/

An example log(LR) histogram is shown in Figure 2.3 for the r'–band and can be compared to that obtained by Rutledge et al. (2000), shown in the top window of Figure 2.1. The Rutledge et al. histogram exhibits a low likelihood tail rising to a maximum at moderate likelihood. The majority of objects lie in this portion of the histogram and are unrelated to the source population. At high LR there are few objects, significantly more of these are situated in the on-source fields than in the background fields, these objects correspond to the true IDs for the sources. The complementary reliability histogram (bottom window of Figure 2.1) shows, as would be expected, that the objects with larger LR values are the most reliable associations and it is easy to see that reliability is proportional to the difference between the on-source and background field LR histograms.

In comparison with the Rutledge et al. LR histogram Figure 2.3 is relatively flat and does not exhibit the tail to low likelihood or the marked difference between the on-source and background field histograms at high LR. This lack of distinction between the histograms manifests itself in large negative spikes in the reliability histogram shown in Figure 2.4. Errors on the reliability were calculated and are shown in Figure 2.5, however, they cannot account for the negative spikes and this is investigated further in

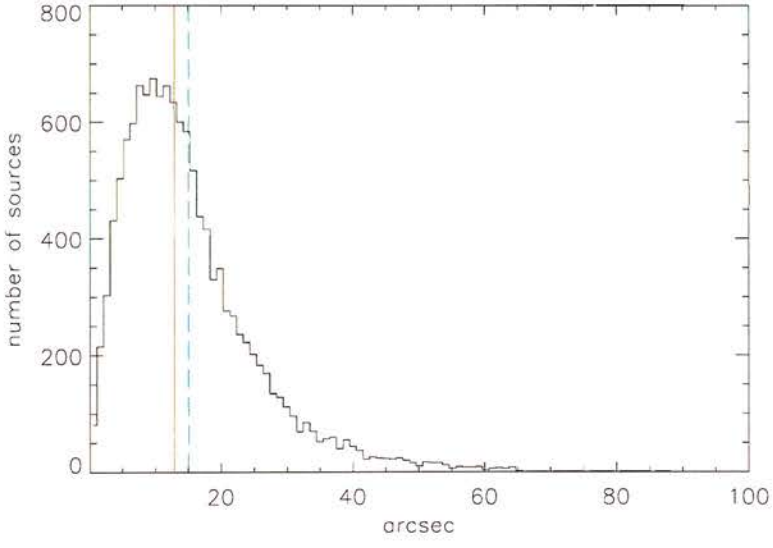

Figure 2.2: Histogram of the distances between simulated source positions and their recovered positions in the simulated maps of Dole et al. (2001). The solid line corresponds to the median at 13 arcsec andthe dashed line is at 15 arcsec.

Section 2.7.1. The low likelihood tail is created by faint objects far from the source, but the search radius used is not large enough to show a significant portion of this. The on-source histogram also exhibits a large peak at moderate likelihood, investigated in Section 2.7.1.

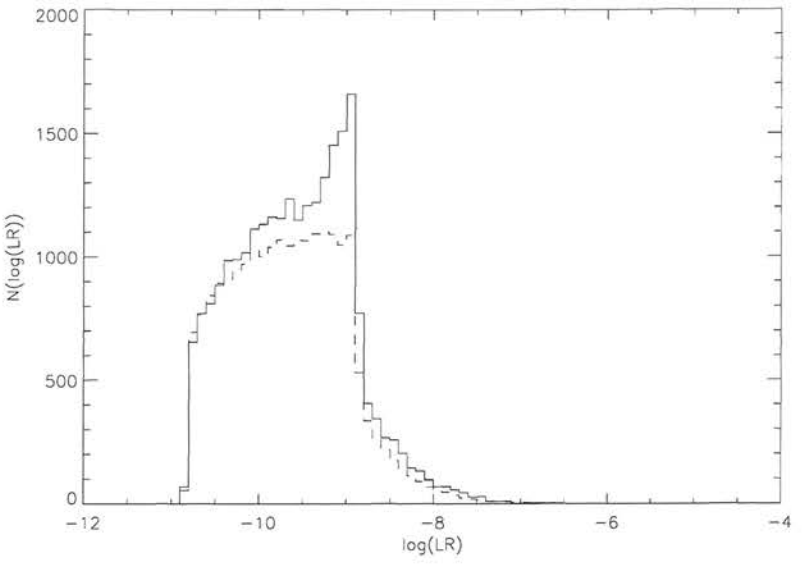

Figure 2.3: A histogram of the log(likelihood ratios) for INT-WFS objects with N1 FIRBACK sources in the r'–band. The on–source fields histogram is shown as a solid line and the background fields histogram as a dashed line.

Likelihood Ratio Histogram Investigations

The likelihood ratios were recomputed using a narrower Gaussian with $\sigma = 20$ arcsec which it was thought might help to reduce the number of objects being assigned moderate likelihoods. However, the excess of objects at moderate likelihood in the on–source fields was still apparent.

Since $LR \propto N(< m)^{-1}$ the magnitude histograms for the on-source and background fields were examined to ascertain whether objects near the far-IR sources are unusually bright which could lead to the LR histogram peak. Figure 2.6 shows an example for the i' band (using $\sigma = 20$ arcsec). The on-source field appears to contain an excess of bright objects which, it was suggested, are caused by the image analyser breaking up large galaxies into multiple sources. The image analyser is unable to handle large galaxies since its de-blending algorithm is too simplistic.

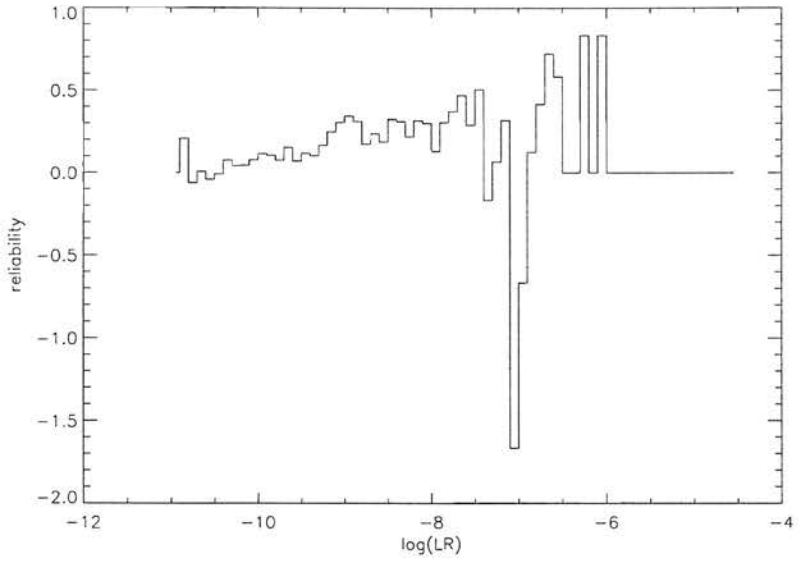

Figure 2.4: A histogram of the reliability of identification in the r'–band as a function of log(LR)

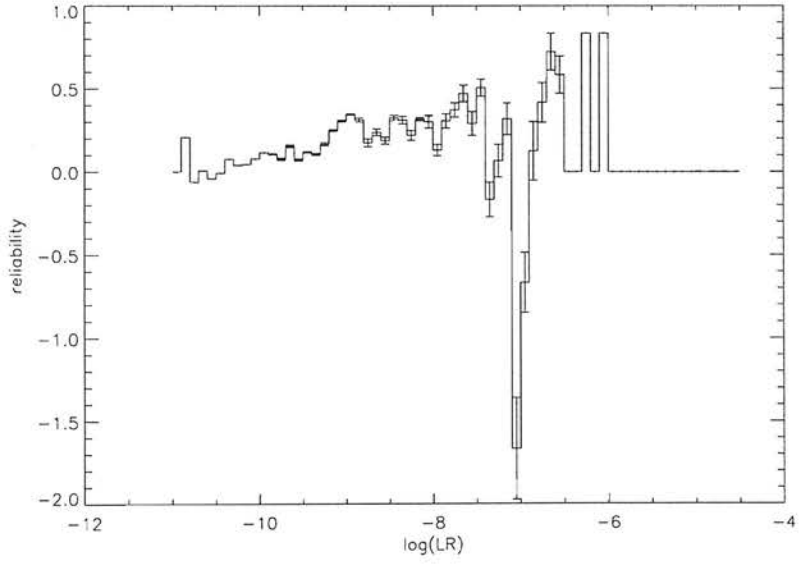

Figure 2.5: Errors on the r'–band reliability.

The most marked likelihood spikes were found in the i' and U–bands, that for the i'–band is shown in Figure 2.7. Objects lying in the peak were marked on images of the on-source fields and those found to lie over the surface of an extended object were noted. An example is shown in Figure 2.8.

Likelihood histograms were constructed separately for those on-source fields containing extended objects and those without. Figures 2.9 and 2.10 show those for objects selected in the i'–band. It is evident that in the histogram for fields that do not contain extended objects the peak has been removed and in that for fields containing all the extended objects the peak is large. We therefore conclude that the spike in the on-source field histograms, which is not correspondingly present in the background fields plots, is due to the image analyser identifying more than one object for extended objects.

Reliability of Identification Investigations

A model was constructed to calculate the probability that the reliability is negative in a particular log(LR) histogram bin. It was modeled that an on-source field contains one extra optical object, compared to a background field, which is the real ID. Since averaged over many background fields it can be assumed that the number of objects

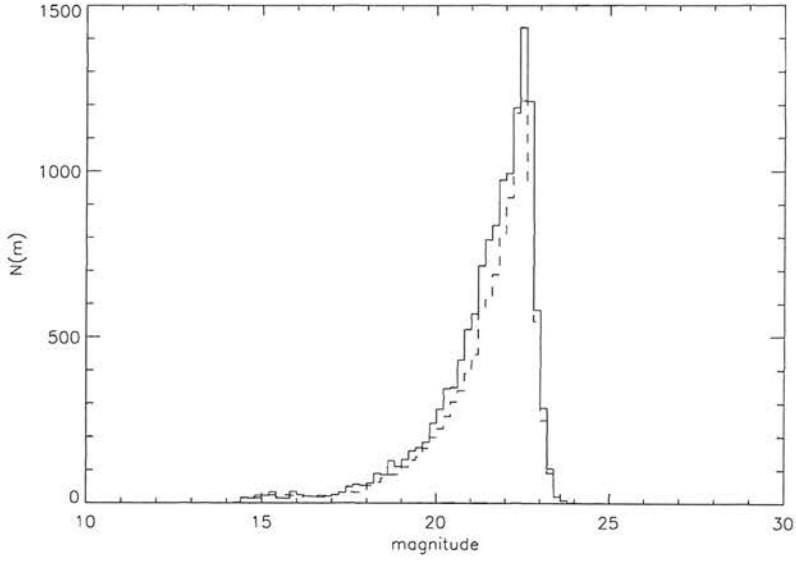

Figure 2.6: The magnitude distribution for the INT-WFS objects in the i'–band. The histogram for on source fields is shown as a solid line and the background fields as a dashed line

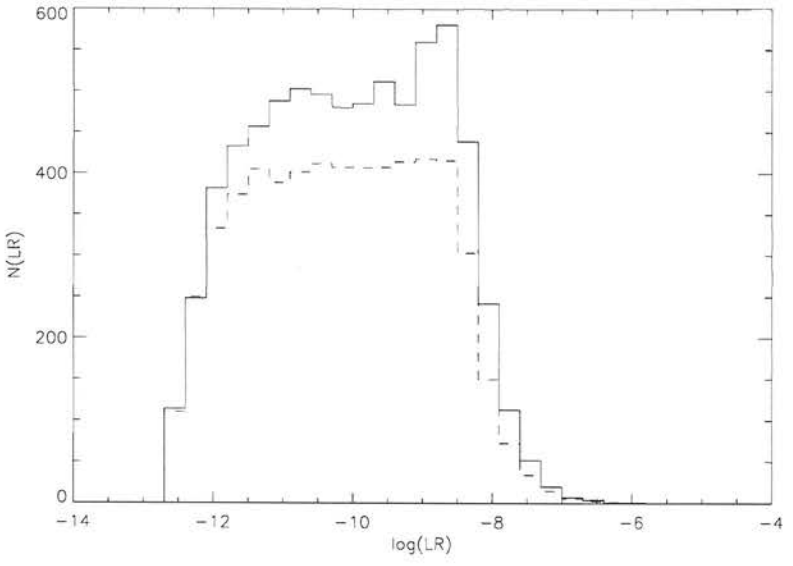

Figure 2.7: Log(likelihood ratio) histograms for objects in the i'–band (σ = 20 arcsec). On–source fields histogram shown as a solid line and background fields as a dashed line.

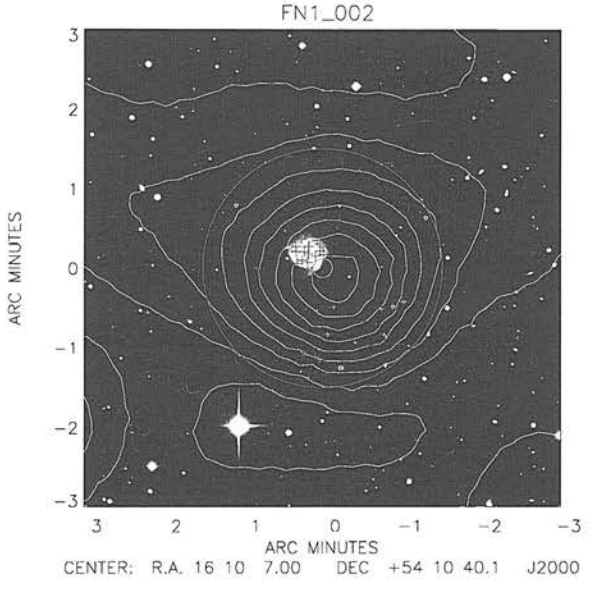

Figure 2.8: An example image for source FN1 002. The source position is marked by the small orange circle and the radius within which counterparts were searched for by the large green circle. Objects with likelihoods lying in the U-band spike are coloured red, there are many objects marked over the face of the galaxy, other objects are marked in blue.

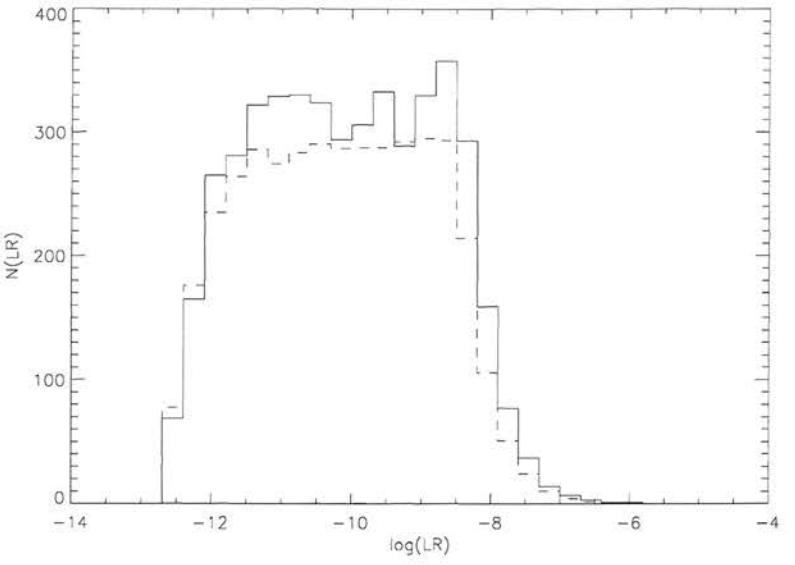

Figure 2.9: The log(likelihood ratio) histogram for source fields, selected in the i'–band, which do not contain extended sources. On–source fields histogram shown as a solid line and background fields as a dashed line.

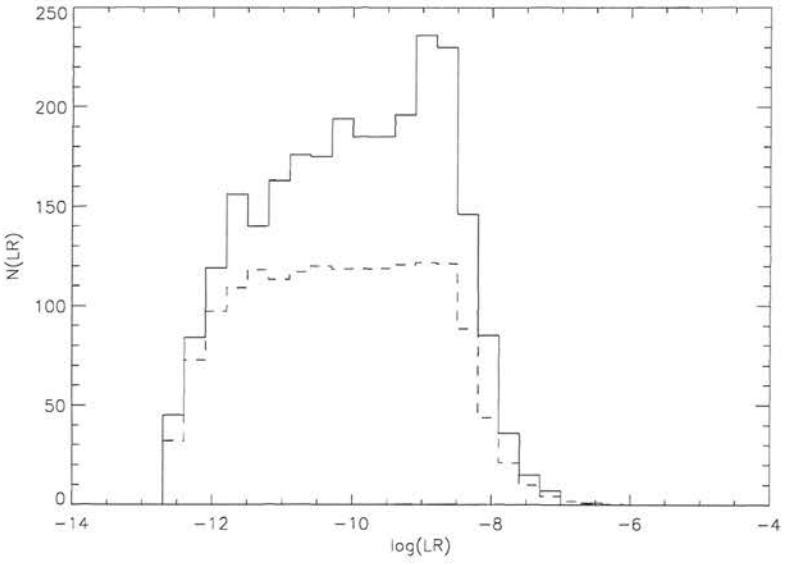

Figure 2.10: The log(likelihood ratio) histogram for source fields, selected in the i'–band, which contain extended sources. On–source fields histogram shown as a solid line and background fields as a dashed line.

per field in background fields, as a function of LR, is constant.

$$N_{background}(LR) = \overline{N}(LR) = const \tag{2.16}$$

The Number of objects per field in the on–source fields as a function of LR is then:

$$N_{source}(LR) = \overline{N}(LR) + prob(LR) \times N_{iso} \tag{2.17}$$

where $prob(LR)$ is the probability that the real ID falls into that LR bin and N_{iso} is the number of ISO FIRBACK sources.

Since reliability is estimated to be (equation 2.11),

$$\widetilde{R}(LR) = \frac{N_{source}(LR) - N_{background}(LR)}{N_{source}(LR)} \tag{2.18}$$

the probability that this is negative in a particular LR bin is equivalent to,

$$P\{(N_{source}(LR) - N_{background}(LR)) < 0\} = \tag{2.19}$$

$$P\{(\overline{N}(LR) + prob(LR) \times N_{iso} - \overline{N}(LR)) < 0\} \tag{2.20}$$

The probability distribution $prob(LR)$ was estimated by constructing a LR distribution for 10,000 fictitious objects. These were modeled to have Gaussianly distributed positional offsets, r, and a top hat distribution of magnitudes.

$\overline{N}(LR)$ was estimated by constructing a LR histogram using the magnitude distribution for the optical catalogue and their positional offsets fitted to an r^2 distribution, since the number of objects surrounding a point on the sky increases as r^2.

$N_{source}(LR)$ could then be constructed for different values of N_{iso}, the number of ISO sources, to see how this affects the probability that the reliability is negative in a particular LR bin.

It was assumed that each LR bin is independent, therefore for each LR bin we have a Poisson distribution with mean N_S [the number in the $N_{source}(LR)$ bin], this is integrated from zero to N_B [the number in the $N_{background}(LR)$ bin] to give the probability that N_S is less than N_B i.e. the probability that the reliability is negative in than particular LR bin. This was repeated for various values of N_{iso}, examples with $N_{iso} = 100$, 1000 and 5000 are given in Figures 2.11 to 2.13.

The probability that the reliability is negative is high for small numbers of sources, it is not until a threshold of about 1000 sources is reached that the probability becomes negligible and therefore the Rutledge et al. (2000) method of reliability calculation becomes applicable. Since there are only 103 N1 FIRBACK sources the probability of the reliability being negative may be as high as 40% in some LR bins, therefore, in this case the Rutledge et al. association technique is not suitable.

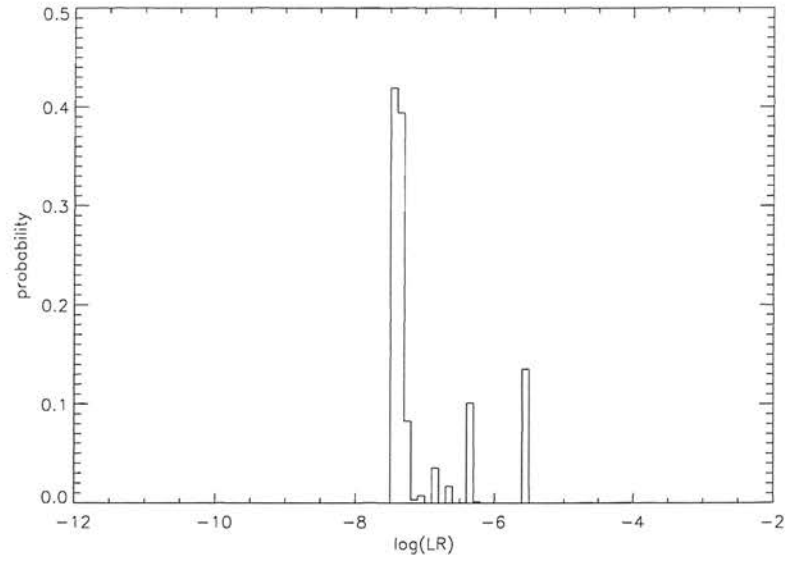

Figure 2.11: The probability that the reliability of association is negative in each log(LR) bin for $N_{iso} = 100$

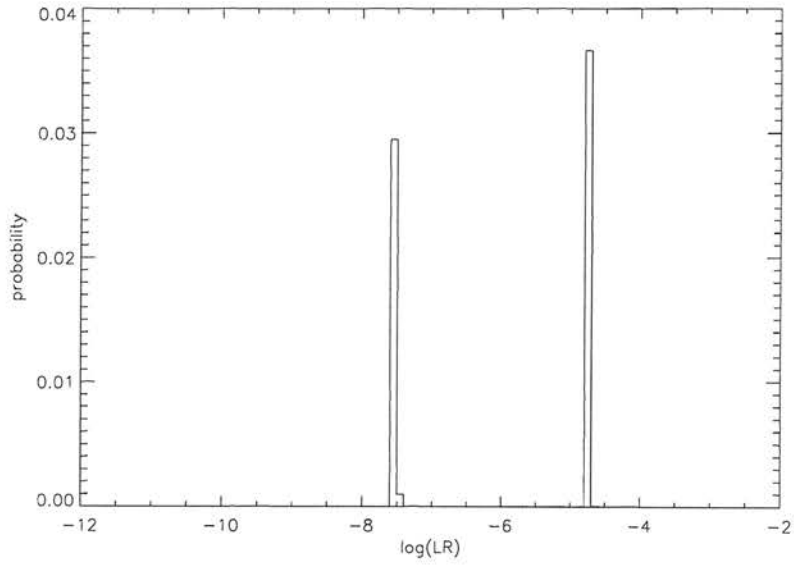

Figure 2.12: The probability that the reliability of association is negative in each log(LR) bin for $N_{iso} = 1000$

2.7.2 The Mann et al. (1997, 2002) Method

Likelihood ratios were calculated for objects within 90 arcsec of each FIRBACK source possessing magnitudes above the completeness limit using an error on the source positions, σ, of 20 arcsec. Figure 2.14 shows the calculated LR histogram in the r'-band. This method has the advantage over the previous one that it is not severely affected by the image analyser identifying spurious objects since $LR \propto 1/N(m)$, therefore, LR is only affected by the small number of extra objects in the object in question's magnitude histogram bin. Whereas, using the Rutledge et al. (2000) prescription $LR \propto 1/N(< m)$, therefore, the likelihood is affected by the cumulative total of extra objects up to the magnitude of the object in question. This is evident in the likelihood histogram (Figure 2.14) as it does not exhibit the spike exhibited in those calculated using the previous method.

As explained in Section 2.6.3, by taking the value of q(f) to be constant, it is not possible to calculate the reliability of identification and instead the probability, P_{ran}, that the LR of the most likely association with a randomly placed source is at least as high as that for the true source is computed. True associations will then have a low value of P_{ran} and a high likelihood ratio. Figure 2.15 shows a histogram of the minimum P_{ran} values for each FIRBACK sources in the N1 field. This can be compared to that

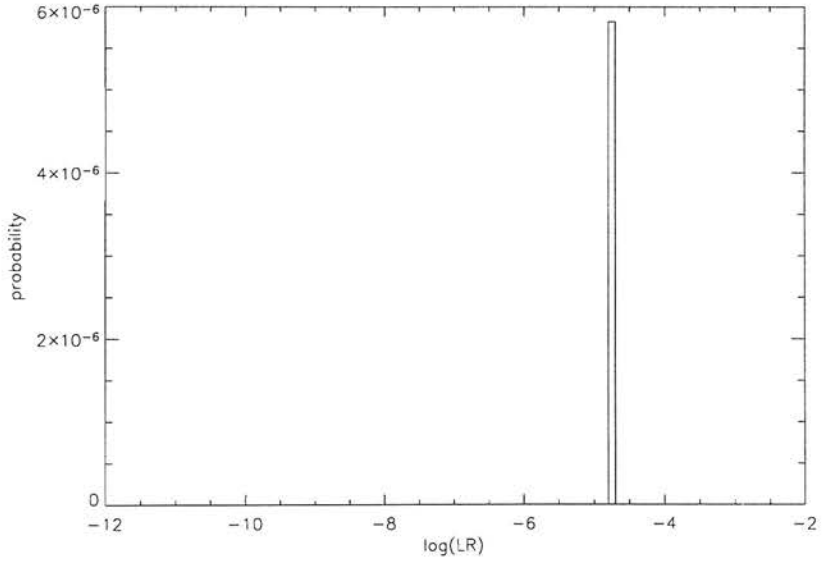

Figure 2.13: The probability that the reliability of association is negative in each log(LR) bin for $N_{iso} = 5000$

found by Mann et al. (2002) for their source list, shown in Figure 2.16. The Mann et al. histogram show that most of their ISO sources have been associated with objects with low P_{ran} indicating that these have a high probability of being the correct ID, these were detected with ISOCAM (Cesarsky et al. 1996) and have small error circles ($\sim$ 6 arcsec). The P_{ran} value taken as the threshold below which association are defined to be good is chosen somewhat arbitrarily through inspection of the P_{ran} histogram. The cut-off for good associations chosen by Mann et al. (2002) is clearly marked ($P_{ran} = 0.15$) on Figure 2.16 and easily made. In the case of the FIRBACK sources this is a little more tricky since the histogram is rather noisy due to the much larger error on the source positions and the fact that the source list may contain spurious sources from the supplementary catalogue, we suggest a threshold of $P_{ran} = 0.4$.

The LR calculations were also carried out using a source positional error, σ, of 45 arcsec. The corresponding likelihood histogram and minimum P_{ran} histogram can be seen in Figures 2.17 and 2.18. In this case the likelihood histogram takes the shape of a reversed magnitude histogram since the dependence of LR on the distance between the source and object has been removed by using a large value of σ.

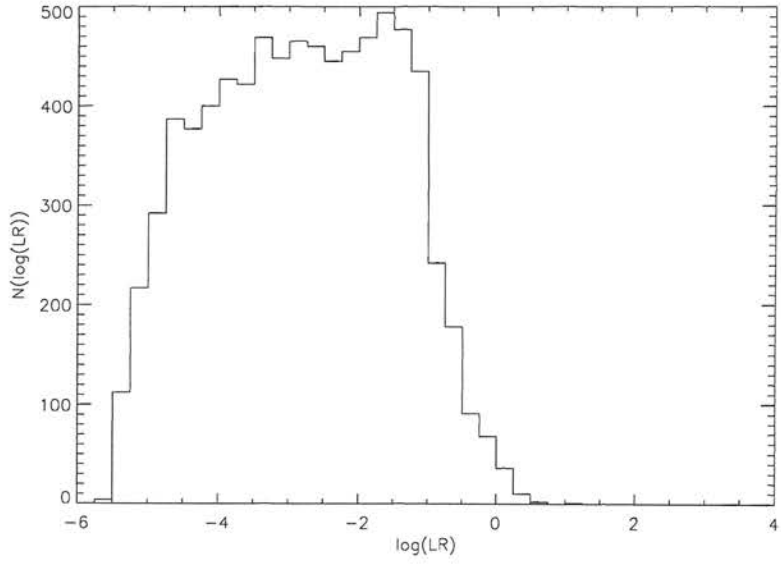

Figure 2.14: The log(likelihood ratio) histogram for FIRBACK sources in the N1 field in the r'–band calculated using the Mann et al. (1997, 2002) method with $\sigma = 20$ arcsec.

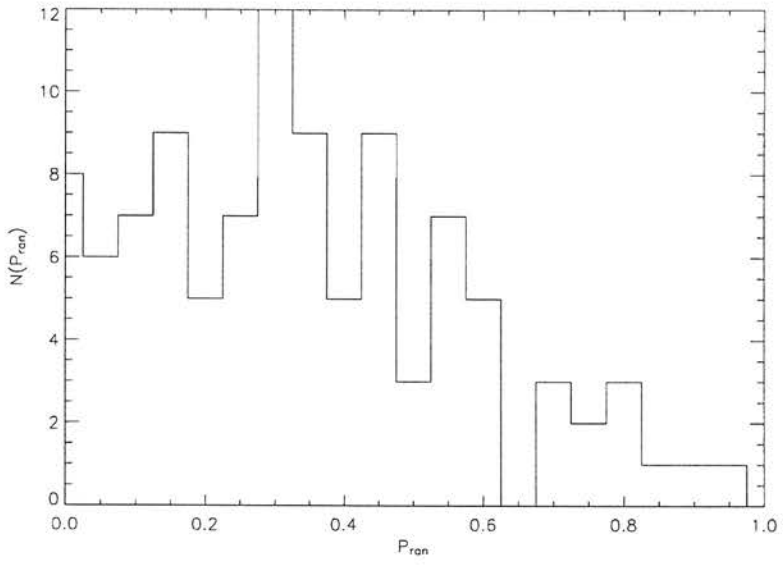

Figure 2.15: The r'–band N1 FIRBACK sources minimum P_{ran} histogram with $\sigma = 20$ arcsec.

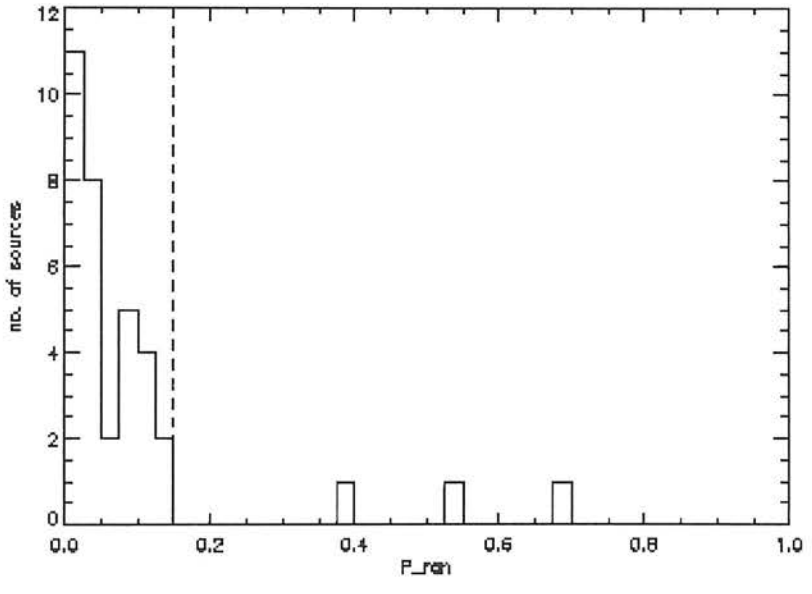

Figure 2.16: The P_{ran} histogram for the Mann et al. ISO sources associated with optical objects. Good associations are defined to be those with $P_{ran} < 0.15$, to the left of the dashed line.

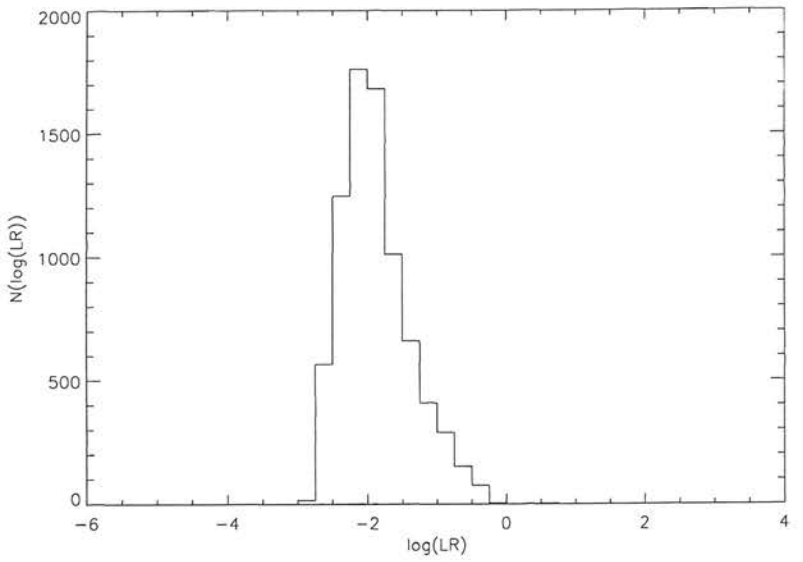

Figure 2.17: The log(likelihood ratio) histogram for FIRBACK sources in the N1 field in the r'–band calculated using the Mann et al. method with $\sigma = 45$ arcsec.

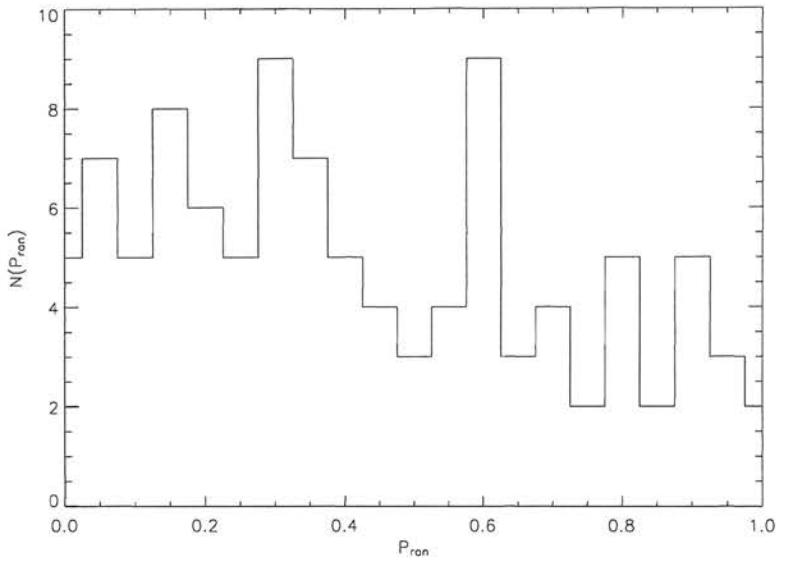

Figure 2.18: The r'–band N1 FIRBACK sources minimum P_{ran} histogram with $\sigma = 45$ arcsec.

Making IDs

Optical IDs for the FIRBACK sources in the N1 field were made by selecting the object with the lowest P_{ran} value for each source i.e. we select the optical object for which it is least likely that a randomly placed object has an LR at least as high as the object in question. The Mann et al. (1997, 2002) likelihood ratio prescription was also applied to find associations for FIRBACK sources in the N2 ELAIS field. This field contains only 55 sources, one of which was discounted as it lies between two WFC chips. The likelihood histograms and minimum P_{ran} histograms using $\sigma = 20$ arcsec and $\sigma = 45$ arcsec are shown in Figures 2.19 to 2.22. From the minimum P_{ran} histograms we again suggest a possible cut-off value of approximately $P_{ran} = 0.4$ for confident associations.

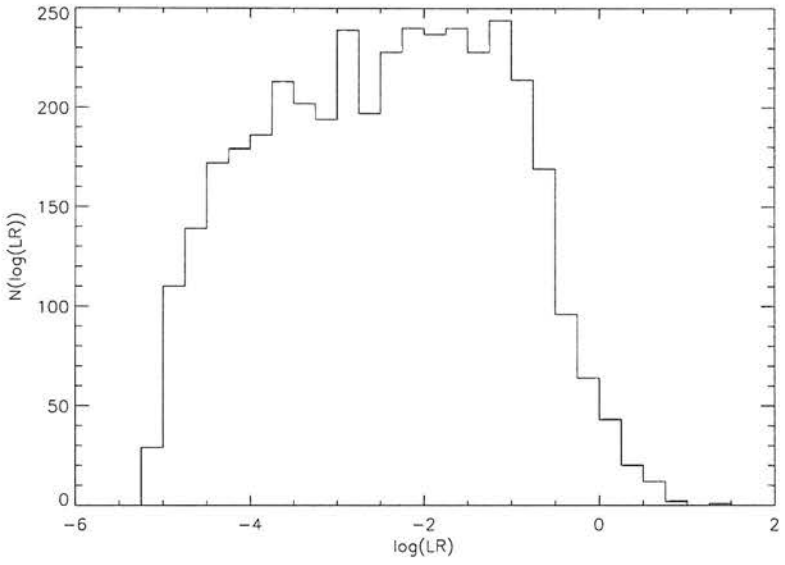

Figure 2.19: The log(likelihood ratio) histogram for FIRBACK sources in the N2 field in the r'–band calculated using the Mann et al. method with $\sigma = 20$ arcsec.

Using this method with a P_{ran} cut-off of 0.4 it was possible to make IDs for approximately 60 per cent of FIRBACK sources in both the N1 and N2 ELAIS fields when using a positional error of $\sigma = 20$ arcsec and approximately 50 percent using $\sigma = 45$ arcsec.

Since far-IR emission is due to light re-emitted by dust we would expect the majority of the optical FIRBACK source IDs to be galaxies rather than stars. However, in several cases this prescription chooses a star as an ID when there is also a galaxy lying within the sources' error ellipse.

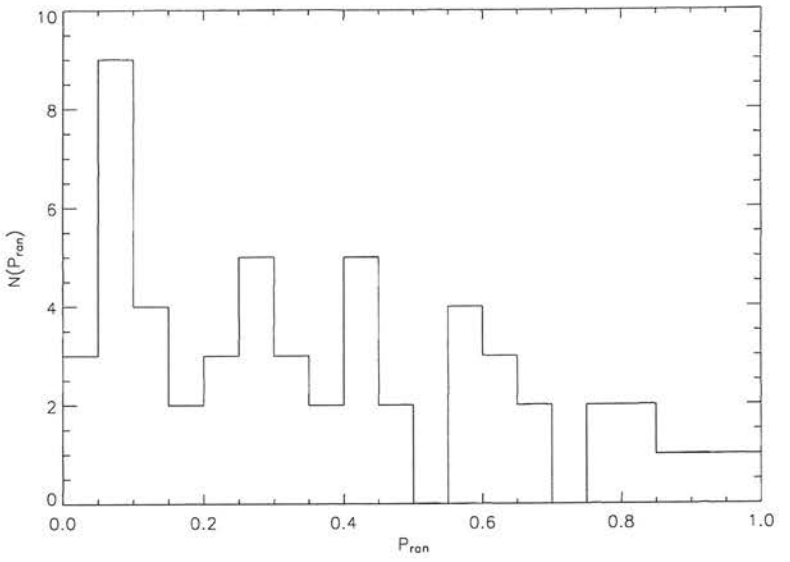

Figure 2.20: The r'–band N2 FIRBACK sources minimum P_{ran} histogram with $\sigma = 20$ arcsec.

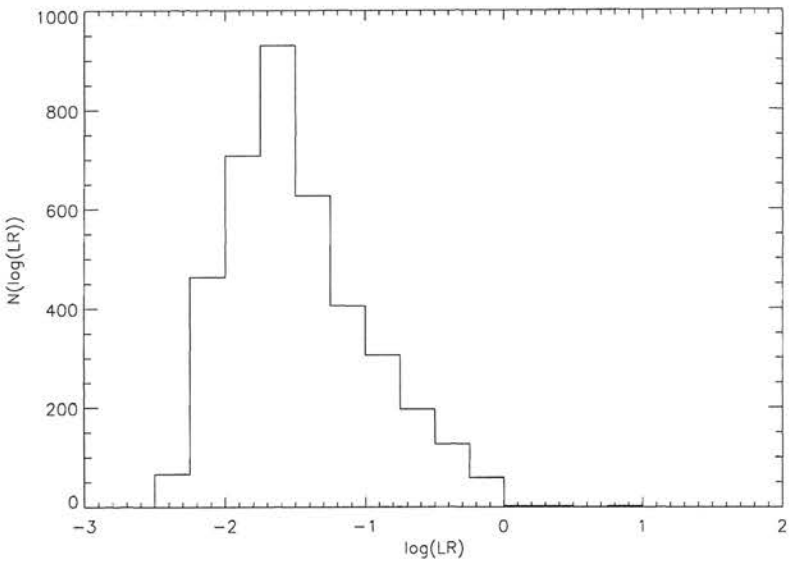

Figure 2.21: The log(likelihood ratio) histogram for FIRBACK sources in the N2 field in the r'–band calculated using the Mann et al. method with $\sigma = 45$ arcsec.

2.8 Summary and Discussion

The likelihood ratio source association method has been tested via searching for INT–WFS optical counterparts for the 175 μm FIRBACK source population. It was discovered that the Rutledge et al. (2000) prescription was not appropriate for small source samples. In this method calculation of the probability that an object is uniquely associated with a source, $P_{id,i}$, to select the most probable ID relies on the calculated values of the reliability of the association with candidate objects within the source error ellipse.

$$P_{id,i} = \left[\frac{R_i}{1 - R_i} \prod_{j=1}^{M} (1 - R_j) \right] / S \tag{2.21}$$

Until a sample size of about 1000 sources is reached the reliability values have a high probability of taking a negative value which renders $P_{id,i}$ meaningless.

The Rutledge et al. (2000) prescription was also affected by extra objects listed in the optical catalogue which had been extracted by the image analyser over extended objects. Since $LR_{\rm Rutledge} \propto 1/N(< m)$ an excess of objects were assigned moderate LR values due to the cumulative inclusion of spurious objects with magnitudes less than that of the object in question. This presented itself as a spike in the on-source field LR histograms, see Figure 2.3.

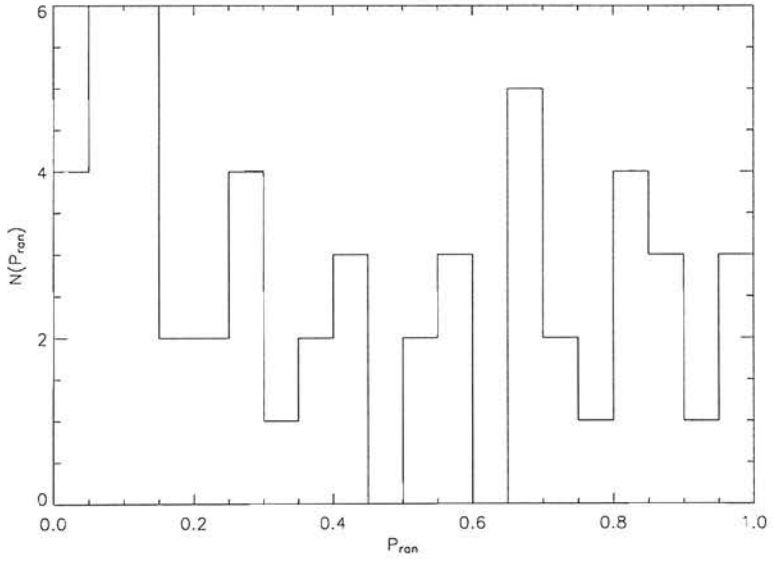

Figure 2.22: The r'–band N2 FIRBACK sources minimum P_{ran} histogram with $\sigma = 45$ arcsec.

However, the Mann et al. (1997, 2002) prescription did not exhibit either of these problems. $LR_{\rm Mann} \propto 1/N(m)$ and is therefore only affected by the small number of spurious objects in that particular magnitude histogram bin. Using this method we were able to successfully assign optical identification to the to the 175 μm sources, selecting those with the lowest P_{ran}, the probability that the LR of the most likely association with a randomly placed source is at least as high as that for the true source, value. Reliable associations were those with P_{ran} less than the threshold value of 0.4. This cut–off must be chosen somewhat arbitrarily from the P_{ran} histogram, e.g. Figure 2.20. This allowed confident ID matches with 60 per cent of the far-IR sources when using $\sigma = 20$ arcsec and 50 per cent of the sources using $\sigma = 45$ arcsec.

In conclusion the Mann et al. (1997, 2002) likelihood ratio prescription was found to be preferable in a case such as this with a small source sample size and where the optical images contain many extended objects from which an image analyser identifies numerous sources. However, in this method it is necessary to choose a P_{ran} threshold value for confident associations rather arbitrarily from the sources P_{ran} histogram. In this particular application this selection was difficult as the histogram was rather noisy (e.g. Figure 2.22) due to the large source positional errors and the fact that the source list may contain spurious sources from the supplementary catalogue. Where positional errors are smaller and the existence of sources more confident the selection of the P_{ran} threshold can be made easily, see Figure 2.16.

The Rutledge et al. (2000) likelihood ratio prescription may be reliably applied when associating over 1000 sources with counterparts from a different catalogue. In this method the probability that an object is the unique association for the source, $P_{id,i}$, for each possible counterpart is calculated as well as P_{no-id}, the probability that the source has no counterpart in the catalogue. This, therefore, gives a quantified threshold for the source having no association. Although the P_{ran} cut–off was not made easily in the Mann et al. (1997, 2002) method IDs were assigned successfully to the majority of the FIRBACK sources. The Rutledge et al. (2000) method was also found to be applicable to large datasets. The likelihood ratio method, therefore, has been proved applicable to challenging association problems where source errors are large. This technique provides an algorithmic encapsulation of human judgement.

2.9 Associating the N2 FIRBACK Optical IDs with further multiwavelength data

The nature of the 175 μm FIRBACK sources was investigated through the association of these sources not only with optical IDs but also with detections at up to nine other wavelengths. This was led by myself in the N2 field (Taylor et al. 2005) and Michel Dennefeld in the N1 field (Dennefeld et al. 2005). The association of the N2 sources with the multiwavelength data is described below.

The N2 FIRBACK sources were first associated with optical objects from the INT–WFS catalogue. This was done using the Mann et al. (1997, 2002) likelihood ratio technique, described earlier. Since, as mentioned previously, far-infrared emission is due to light re-emitted by dust, it would expected that the FIRBACK source IDs are galaxies. Therefore, when the likelihood ratio method was applied to objects within 90 arcsec of the sources, those classified as stars were forced to have $P_{ran} = 1.0$ to avoid the selection of a stellar ID. The width of the Gaussian distribution of positional offsets between the objects and the sources was taken as $\sigma = 20$ arcsec. This prevents too many objects being assigned moderate likelihoods, without excluding objects which should be considered or including objects which should not and also ensures that the likelihood ratio is still sensitive to source–counterpart separation. The threshold value for acceptable IDs was chosen to be $P_{ran} = 0.15$ through inspection of the distribution of minimum P_{ran} values, see Figure 2.23.

Associations were then also sought with sources in the multiwavelength ELAIS (European Large Area ISO Survey) band-merged catalogue at 6.7 μm, 15 μm and 90 μm with associated data at U,g',r',i',Z,J,H,K and 20 cm (Rowan-Robinson et al. 2004). This was a survey carried out using the Infrared Space Observatory (ISO) (Kessler et al. 1996) covering a total of 12 deg^2 over five main fields, three in the north (N1, N2, N3) and two in the south (S1,S2). These areas were also surveyed at 20 cm with the VLA and AT. The separate wavebands making up the ELAIS survey each comprise an independent survey, the band–merged catalogue simply bringing together the results of each of their final analyses. The multi-wavelength ELAIS catalogues were merged, taking into account the different positional accuracies of the component catalogues, sequentially as follows. The 15 μm and 20 cm catalogues (Vaccari et al. 2005; Ciliegi et al. 1999) were first identified with optical objects from the INT–WFS catalogue (González–Solares et al. 2005) and then merged according to their optical positions using a search radius of 2 arcsec. These were then matched with the 6.7 μm catalogue

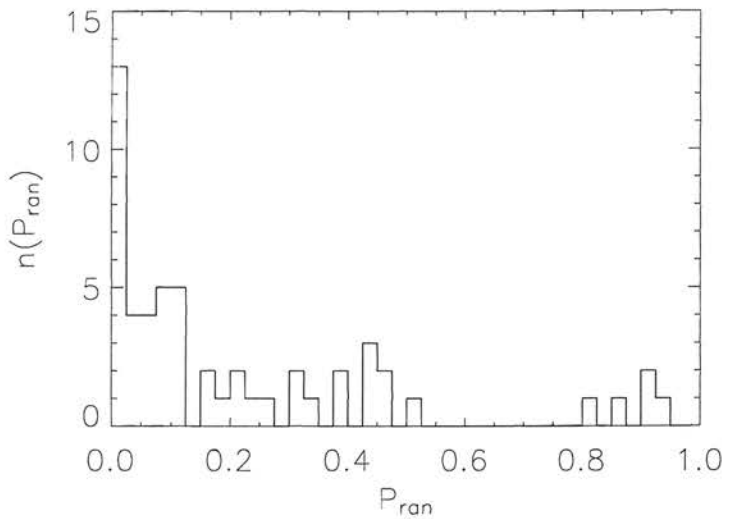

Figure 2.23: Distribution of P_{ran} values for chosen source counterparts, selected in the r'–band. Galaxies only, $\sigma = 20$ arcsec.

using a search radius of 5 arcsec and those 6.7 μm sources not merged were associated with optical counterparts. These were then merged with the 90 μm and 175 μm sources (Héraudeau et al. 2004; Dole et al. 2001) using search radii of 30 arcsec and 60 arcsec respectively. Where these were matched with more than one catalogue entry the less probable associations were flagged in the catalogue. After this those sources that are still unmatched were identified with optical counterparts.

To associate our 175 μm sources with the ELAIS catalogue we used a search radius of 60 arcsec in consistency with the ELAIS band merging process. As described above optical counterparts were sought by Rowan-Robinson et al. (2004) independently from this work using the FIRBACK catalogue and the resulting optical IDs were then compared.

In summary, we sought associations on the basis of applying the likelihood ratio method to the optical data alone, and then, when the final band-merged ELAIS catalogue became available, we reassessed them. Our final optical and ELAIS wavelength ID list was drawn up using the following criteria:

1. If our optical ID, selected using the likelihood ratio method, agreed with the ELAIS optical ID to which the 175 μm flux had been assigned then both the ELAIS optical ID and band-merged data were accepted.

2. If the ELAIS optical ID had a $P_{ran} < 0.15$ (cut off value see earlier) for association with the 175 μm ISO source then the ELAIS optical ID and band-merged data were accepted.

3. If our optical ID and ELAIS optical ID were in fact the same galaxy but have different entries in the INT–WFS catalogue due to the way the image analyser splits up bright galaxies into multiple sources, then the band-merged ELAIS data and optical ID were accepted.

4. If our optical ID had $P_{ran} < 0.15$ but no optical ELAIS catalogue match then the source has optical and 175 μm data only (it may also have IRAS detections/upper limits and 60 μm and 90 μm upper limits)

5. If our optical ID had $P_{ran} > 0.15$ the source has no association

Where our sources were not associated with an ELAIS catalogue entry or where the associated entry did not have 15 μm or 90 μm detections, upper limits were extracted directly from the 15 μm and 90 μm survey maps (Vaccari et al. 2005; Héraudeau et al. 2004) at a 3σ level using aperture photometry with an aperture size of 6 arcsec and 90 arcsec respectively centred on the optical positions. σ was determined by calculation of the flux falling on 100 apertures distributed in a regular grid around the source position. Flux totals of more than 3σ in an aperture positioned over the source were taken to be a detection and those with fluxes less than this were given 3σ level upper limits of: flux within the on source aperture $+3\sigma$. An aperture correction was applied to the 15 μm upper limits since 40 per cent of the psf lies outwith the aperture.

60 μm and 100 μm fluxes or upper limits were also obtained for each of the sources via the IRAS (Infrared Astronomical Satellite) SCANPI facility[1]. IRAS conducted a sensitive and unbiased survey of over 96 per cent of the sky at 12 μm, 25 μm, 60 μm and 100 μm. SCANPI, the IRAS interactive service tool, averages the calibrated survey scans in order to return fluxes or upper limits at a given position. We obtained 3σ 100 μm and 60 μm fluxes or upper limits at the sources' optical identification positions

Associations were also sought in the Submillimetre Common User Bolometer Array (SCUBA) 850 μm (Scott et al. 2002) and Max–Planck Millimeter Bolometer array (MAMBO) 1200 μm (Greve et al. 2004) catalogues, however, none were found for any of our 175 μm sources.

[1] http://irsa.ipac.caltech.edu/applications/IRAS/Scanpi/

For sources in the main N2 FIRBACK catalogue ($S > 180$ mJy), which contains 27 detections, we made 22 confident optical associations with galaxies, 7 of which agree exactly with the ELAIS optical ID (category i.) and 6 for which we accept the ELAIS optical ID (categories ii. or iii.). Nine confident optical associations have been made for sources in the supplementary catalogue ($S < 180$ mJy, 28 sources), 2 of which agree with the ELAIS optical ID (category i.) and 2 of which where the ELAIS optical ID is accepted (categories ii. or iii.). There were also two sources in the supplementary catalogue with stellar IDs, one of which agreed with the ELAIS optical ID (category i.) and one of which was an accepted ELAIS optical ID (category ii. or iii.). The data for the identified sources are given in Table 3.1 in Chapter 3 and notes on individual sources are given in Appendix A including P_{ran} values of any other plausible identifications.

To summarise, the likelihood ratio method made confident optical associations with galaxies for 31 out of 54 far–infrared sources. The selection of 8 of these was altered by the inclusion of further multiwavelength data. These data can now be used to investigate the nature of the FIRBACK sources (see Chapter 3) thus illustrating the value of associating detections from different catalogues at different wavelengths and the need for reliable association algorithm techniques to make this possible.

It is much easier to make measurements than to know exactly what you are measuring.

J. W. N. Sullivan 1886-1937

3

The Nature of the 175μm Selected N2 FIRBACK Sources

3.1 Introduction

This chapter illustrates the importance of associating sources from different surveys at different wavelengths. We have gone one step further than previous investigations into the nature of the FIRBACK sources by seeking associations for all 55 N2 175 μm ISO sources in the FIRBACK survey not only with optical sources from the Isaac Newton Telescope – Wide Field Survey (INT-WFS) catalogues in four bands (g',r',i',Z), but also with the other ELAIS survey wavelengths (J, H, K, 15 μm, 6.7 μm, 90 μm, 20 cm) and IRAS wavelengths (60 μm, 100 μm) (described in Chapter 2). In doing so we are able to gain a multiwavelength view of the nature of these interesting objects. We have compared their spectral energy distributions (SEDs) with those predicted by the radiative transfer models of Efstathiou and Rowan-Robinson (2003) to provide further insight into their properties. (These results have been published in Taylor et al. 2005, given in Appendix B). We also investigated sources in the ELAIS N1 region (published in Dennefeld et al. 2005, also in Appendix B), a summary of whose results is given in

Section 3.8.

The layout of this chapter is as follows: Section 3.2 briefly describes the optical properties of the sources including their morphologies and redshift distribution. The radiative transfer models are described in Section 3.3 and Section 3.4 describes the comparison of the sources' SEDs with predictions from the models. The resulting star formation rates, far-infrared luminosities and dust temperature estimations are given in Sections 3.5 and 3.6. Finally, a discussion of the nature of the far-infrared background sources and comparison of results with current evolutionary model predictions is given in Section 3.7.

3.2 Optical Properties

The associated ELAIS and IRAS multiwavelength data for the identified sources are given in Table 3.1 and notes on individual sources are given in Appendix A, including P_{ran} values for any other plausible identifications.

The identified sources were morphologically classified by Simona Mei by visual inspection. She used i'-band images from the INT–WFS, and followed the same classification method as Postman et al. (2005). The galaxies were classified using the classical Hubble sequence: E, E/S0, S0, S0/a, Sa, Sa/b, Sb, Sb/c, Sc, Irr. For the Advanced Camera for Surveys on HST, for galaxies with redshift of around unity, visual classifications of this kind have a typical random error of 25 per cent, which reduces to 6 per cent when galaxies are classified in two broad categories: 1. early-type, 2. spirals and irregulars. These errors were estimated from the average scatter found in classifications by four ACS team members. The classification will be less reliable for our data due to poorer angular resolution in the INT data caused by atmospheric seeing. The majority of the sources are spiral galaxies, there are also a number of S0 galaxies, 2 sets of interacting galaxies, 2 pairs and 2 irregulars, the morphology of 6 sources was unable to be deduced from the images.

Table 3.1: Multi-wavelength data for associated sources, fluxes in mJy. N.B. Sources listed P_{ran} greater than threshold 0.15 due to falling into category (iii) (see Section 2.9), FN2 004 has $P_{ran} = 1.0$ since the associated galaxy is mistakenly classified as a star in the INT WFS catalogue.

FIRBACK	FIRBACK position	optical position	175 μm	P_{ran}	r'-mag	15 μm	6.7 μm	60 μm	90 μm	100 μm	20 cm	z_{phot}	z_{spec}	morphological classification
FN2 000	16 37 33 +40 52 26	16 37 34.53 +40 52 11.2	2377±213	0.026	18.46	53.9±0.3	18.3±0.1	-	1461±36	-	8.74±0.02	-	0.03	irregular or edge–on spiral with tidal tail
FN2 001	16 35 08 +40 59 20	16 35 07.87 +40 59 28.9	1251±139.	0.002	18.28	26.1±0.4	4.11±0.12	0.15±0.03	337±15	-	0.76±0.04	0.10	0.03	edge–on Sb/c
FN2 002	16 36 10 +41 05 16	16 36 08.15 +41 05 07.7	803±102	0.086	19.33	8.9±0.1	2.3±0.1	0.41±0.02	614±37	-	-	0.15	0.17	irregular with tidal tail
FN2 003	16 35 25 +40 55 51	16 35 25.22 +40 55 42.1	682±92	0.001	17.45	13.9±0.1	5.5±0.1	0.19±0.03	416±19	-	1.70±0.02	0.17	0.03	Sa
FN2 004	16 34 01 +41 20 49	16 34 01.82 +41 20 52.5	666±91	1.000	16.41	20.4±0.1	9.6±0.1	0.40±0.02	403±26	-	1.8±0.02	0.05	0.03	SBa
FN2 005	16 32 43 +41 08 38	16 32 42.39 +41 08 46.1	522±78	0.011	18.67	6.1±0.2	2.1±0.1	0.16±0.03	399±27	-	0.67±0.04	0.02	0.26	Sa
FN2 007	16 35 45 +40 39 14	16 35 46.91 +40 39 03.4	316±60	0.022	17.95	6.1±0.2	2.4±0.1	0.09±0.03	84±25	-	0.92±0.02	0.02	0.12	compact source with 2 nuclei
FN2 008	16 35 47 +41 28 58	16 35 48.04 +41 28 30.3	293±58	0.084	18.32	4.5±0.1	1.3±0.1	0.12±0.03	74±18	-	0.38±0.01	0.10	0.14	Sb
FN2 010	16 35 38 +41 16 58	16 35 36.16 +41 17 27.3	285±57	0.451	19.47	2.8±0.6	-	0.06±0.03	74.00±17	-	0.80±0.01	0.02	0.17	pair: irregular with tidal tail and Sa
FN2 011	16 38 07 +40 58 12	16 38 08.78 +40 58 07.4	260±55	0.118	19.78	<1.97	-	<0.12	<0.44	<0.26	-	0.51	-	Sa?
FN2 012	16 34 13 +40 56 45	16 34 11.98 +40 56 52.8	249±54	0.088	19.99	2.7±0.1	1.0±0.1	<0.08	83±10	-	0.51±0.01	0.07	0.14	Sa?
FN2 015	16 36 07 +40 55 37	16 36 07.71 +40 55 47.1	223±51	0.015	18.64	5.5±0.1	1.2±0.1	<0.06	109±11	-	0.81±0.02	0.26	0.17	interacting galaxies: Sa/b and Sa
FN2 016	16 34 26 +40 54 07	16 34 23.90 +40 54 10.0	218±51	0.191	19.88	3.1±0.1	1.0±0.1	0.06±0.03	<0.25	0.20±0.08	0.55±0.01	0.10	0.13	edge–on: type not clear
FN2 017	16 34 44 +41 08 42	16 34 44.90 +41 08 20.6	213±50	0.030	18.46	2.2±0.7	-	<0.10	<0.25	<0.18	-	0.18	-	interacting galaxies: Sa with edge–on
FN2 018	16 33 38 +41 01 15	16 33 37.23 +41 01 09.1	212±50	0.037	19.51	<2.98	-	0.08±0.03	<0.28	<0.19	-	0.15	-	S0/a
FN2 019	16 37 17 +40 48 36	16 37 16.80 +40 48 25.6	205±49	0.001	18.05	<3.99	-	0.06±0.04	<0.26	<0.18	-	0.10	0.03	Sb
FN2 020	16 32 41 +41 06 10	16 32 40.50 +41 06 15.4	201±49	0.086	20.21	<6.58	-	<0.07	<0.25	<0.11	-	0.15	-	-
FN2 021	16 37 58 +40 51 21	16 37 59.39 +40 51 15.8	196±49	0.001	17.57	<2.18	-	<0.07	<0.19	<0.17	-	0.10	-	-
FN2 022	16 37 08 +41 28 26	16 37 08.21 +41 28 56.1	190±48	0.140	18.80	2.5±0.1	1.7±0.2	<0.10	<0.28	<0.23	0.56±0.02	0.05	0.17	Sa
FN2 023	16 33 51 +40 49 44	16 33 51.65 +40 49 46.3	188±48	0.018	19.04	2.5±0.7	-	<0.11	<0.28	<0.32	-	0.02	-	Sb/c
FN2 025	16 36 31 +40 47 38	16 36 31.23 +40 47 24.7	184±48	0.119	20.24	2.1±0.7	-	0.11±0.03	<0.33	0.11±0.07	-	0.20	-	Sa
CFN2 029	16 34 20 +41 06 54	16 34 19.46 +41 06 37.8	178±47	0.070	19.62	<4.36	-	0.06±0.02	<0.21	<0.24	-	0.15	-	-
CFN2 030	16 35 23 +40 38 42	16 35 22.81 +40 38 37.1	178±47	0.086	20.26	<3.66	-	0.06±0.03	<0.24	0.17±0.05	-	0.15	-	S0/a
CFN2 034	16 34 12 +40 46 26	16 34 12.52 +40 46 34.4	166±46	0.030	19.42	<2.61	-	<0.08	<0.11	<0.10	-	0.55	-	S0/a
CFN2 036	16 37 01 +40 43 08	16 36 59.92 +40 42 46.6	165±46	0.119	19.40	<5.08	-	0.04±0.03	<0.25	<0.28	-	0.10	-	Sa
CFN2 038	16 34 32 +41 22 37	16 34 31.57 +41 22 45.7	161±45	0.023	19.26	1.8±0.1	-	<0.08	<0.17	0.12±0.05	<0.15	0.02	0.14	Sb/c
CFN2 039	16 36 13 +40 42 25	16 36 13.65 +40 42 30.0	160±45	0.001	17.15	5.9±0.2	2.5±0.1	0.11±0.00004	90±21	-	0.88±0.01	0.02	0.07	Sb
CFN2 044	16 37 26 +40 45 39	16 37 25.93 +40 45 37.1	150±44	0.053	19.96	<2.87	-	0.02±0.05	<0.30	<0.14	-	0.05	-	S0/a
CFN2 047	16 34 51 +41 20 27	16 34 49.54 +41 20 49.2	147±44	0.125	19.16	2.4±0.1	1.3±0.1	<0.05	<0.20	<0.21	0.15±0.04	0.07	0.25	S0/a
CFN2 049	16 37 42 +41 19 11	16 37 41.44 +41 19 14.8	143±44	0.210	20.87	1.9±0.2	-	0.09±0.03	74±24	-	<0.15	0.51	-	Sa?

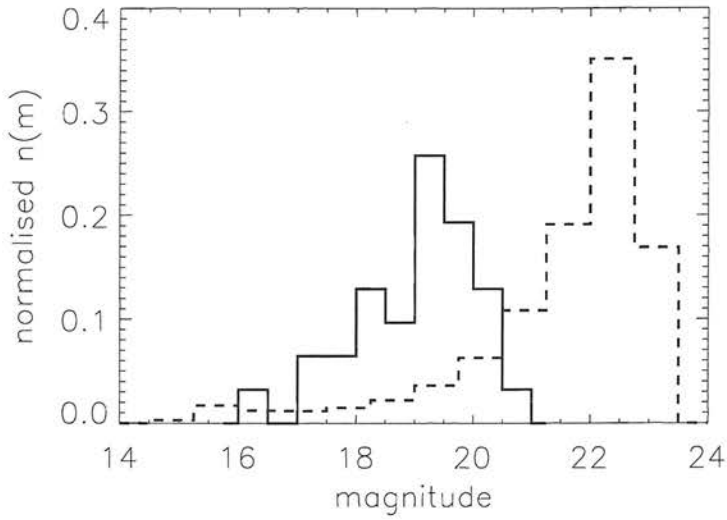

Figure 3.1: Normalised magnitude distribution of the N2 FIRBACK source optical counterparts (r'–band) (solid line), dashed line gives the r-band magnitude distribution of the INT-WFS catalogue for comparison.

Figure 3.1 shows the optical (r'–band) magnitude distribution of the N2 FIRBACK source optical counterparts compared to that of the overall INT-WFS catalogue. The source IDs are, in general, bright objects with apparent magnitudes ranging from 21 to as bright as 16.

Figure 3.2 shows the redshift distribution of the sources. The sources lie predominantly at $z < 0.3$ with only two sources having $z > 0.5$.

3.3 Radiative Transfer Models

It was initially attempted to fit the data to the GRASIL model SEDs of Silva et al. (1998). These provide six templates for nearby galaxies, three starbursts (Arp 220, M82 and NGC 6090) and three spirals (M100, M51 and NGC 6946). However, these templates did not describe the data well. Instead we employed the radiative transfer model codes of Efstathiou et al. (2000) and Efstathiou and Rowan-Robinson (2003) to generate SED predictions for comparison with the data. These are able to model both cirrus and starburst emission or a combination of the two. The models used are briefly described below.

Ultraviolet to submillimetre emission from stars embedded in interstellar dust in

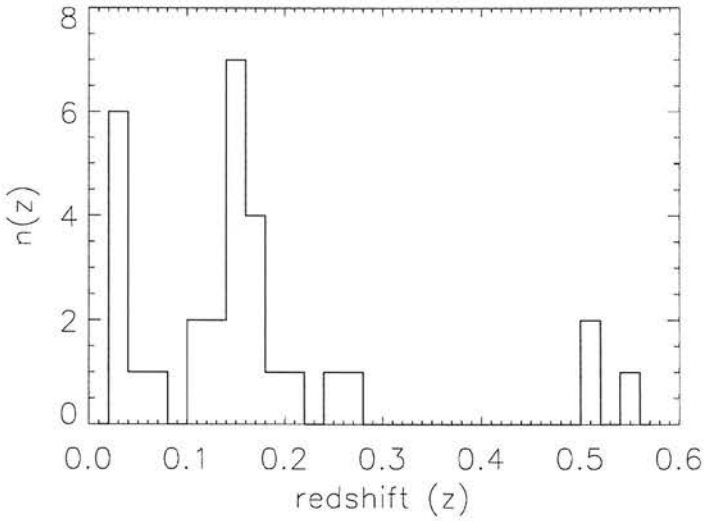

Figure 3.2: Redshift distribution for the 30 175μm sources with spectroscopic or photometric redshifts as listed in Table 1.

galaxies known as the 'infrared cirrus component' is modeled. The model consists of the input stellar radiation field, an interstellar dust model and the radiative transfer treatment of the interaction between the two which generates the output spectral energy distribution.

The input stellar radiation field is taken from the Galaxy Isochrone Synthesis Spectral Evolution Library (GISSEL, Bruzual and Charlot 1993). This gives the radiation from a mass of stars from the ultraviolet to near-infrared as a function of time.

Stars form in molecular clouds and at the early stages of their life remain inside them, during which time their radiation is absorbed by the dust and reprocessed to the infrared. During this phase of their life the model uses the code of Efstathiou et al. (2000) to compute the radiation emitted by stars and dust. The cloud is assumed to disperse 7.2×10^7 years after star formation. Before this, however, due to non-spherical evolution of the cloud, a fraction of the starlight f may have been able to escape without dust absorption. This occurs a time t_m after star formation.

The radiation field within the galaxy is modeled to be due to a large number of randomly oriented molecular clouds whose average emission, for stars in the age range t_m to 7.2×10^7 years, is approximately

$$(1 - f)S_\nu^S + fS_\nu^* \tag{3.1}$$

Table 3.2: Parameter values

	starburst	non-starburst
τ	8 Gyrs	1 Myr
t_{SB}	galaxy age$-t_m$	-
galaxy age	(universe age at z - 0.5Gyrs)	
τ_{SB}	40 Myrs	-
leak	0.0	0.1
f	1.0	1.0
t_m	20 Myrs	3 Myrs
A_V	0–3	0–10
τ_v	50–200	-
ψ	5	0.5-12.5
f_{SB}	1-100	-

were S_ν^S is the emission from a giant spherical molecular cloud (Efstathiou et al. 2000) and S_ν^* is the emission from the stellar population (Bruzual and Charlot 1993). The emission of stars younger than t_m is simply S_ν^S and those older than 7.2×10^7 years S_ν^*. The fraction of general starlight that escapes the galaxy unattenuated is also parameterised in a 'leak' variable.

The dust temperature is determined by the intensity of the stellar radiation field and is characterised in the model by the ratio of the bolometric intensity of the radiation field to that in the stellar radiation field in the solar neighbourhood (Mathis et al. 1983), ψ. The star formation rate, $\dot{\phi}_*(t)$, is assumed to have an exponential form with a time scale, τ, and a Salpeter IMF from 0.1-125 $M_\odot$.

$$\dot{\phi}_*(t) \propto e^{\frac{-t}{\tau}} \tag{3.2}$$

In the case where a starburst is included the star formation rate is modified to

$$\dot{\phi}_*^{SB}(t) \propto \left[\dot{\phi}_*(t) + f_{SB} e^{\frac{-(t-t_{SB})}{\tau_{SB}}}\right] \tag{3.3}$$

where f_{SB} is the ratio of the star formation rate at the peak of the starburst to that at time 0, t_{SB} is the age of the galaxy at the start of the starburst and τ_{SB} is the exponential timescale of the starburst.

The interstellar dust model of Siebenmorgen and Kruegel (1992) is used, which models the effects of both small grains and polycyclic aromatic hydrocarbons (PAHs). The opacity of the dust is characterised by A_V which controls the amount of UV–to–near-IR light absorbed by dust and re-emitted in the IR and submm and therefore the

ratio of the luminosity in these two bands. The model also has the ability to increase the dust extinction of the starburst component of the galaxy separately via the parameter τ_v; values of $50, 100, 150, 200$ are supported. In the starburst case we used the model code to produce a grid of models over f_{SB}, A_V and τ_v and for non-starburst galaxies a grid of models was produced over A_V and ψ. The values of the parameters used are given in Table 3.2.

The effects of each of the model parameters on the resulting SED is shown in Figure 3.3. The solid black line shows an SED produced using the parameters fixed ($\psi = 5$, $\tau = 8$ Gyrs, $t_m = 20$ Myrs, $\tau_{SB} = 40$ Myrs) and $f_{SB} = 1$, $A_V = 1$, and $\tau_v = 100$ for comparison. A model produced with increased dust extinction of $A_V = 3$ is shown by the black dotted line. This increases the emission at longer wavelengths and decreases that in the optical, as would be expected. The dash-dot-dot-dot line shows a model with $t_m = 0$ which dramatically changes the shape of the far-infrared peak due to the reduction in the amount of dust reprocessing of radiation from newly formed stars. However, any significant amount of time spent within a molecular cloud produces a far-infrared peak of shape similar to that of the other models shown. The peak of the SED is also increased significantly with an increase in the strength of the starburst, a model with $f_{SB} = 5$ is shown by the grey dashed line. The other parameters do not affect the resulting SED vastly, hence, there are degeneracies and a spread of parameter values may result in a very similar SED. This is evidenced by models with $\psi = 10$ (dashed line), $\tau = 80$ Gyrs (dash-dot), $\tau_{SB} = 80$ Myrs (long dashes) and $\tau_v = 200$ (solid grey line) all having very similar shaped SEDs. The parameter values chosen were in the mid-range of physically viable values as found by Efstathiou et al. (2000). The τ_v range covers values known to fit normal starbursts ($\tau_v = 50$) (Efstathiou et al. 2000) to heavily obscured starbursts such as Arp 220 ($\tau_v = 200$), this range was also used by Farrah et al. (2003) to fit a number of ULIRGs.

After fitting the models to the multiwavelength data, the parameter values for models within the 1σ χ^2 confidence limit of the best–fitting model were examined. The 1σ limit for starburst model fitted SEDs typically encompasses $\tau_v \pm 50$, $A_V \pm 0.5$ and $f_{SB} \pm 1$. There is also a correlation between f_{SB} and A_V such that models with higher f_{SB} values require low A_V values to fit the data. This is due to an increase in each of the parameters increasing the height of the far–infrared peak (as shown in Figure 3.3), therefore when both the parameters have high values the height of the peak is over–estimated. We also note that it is difficult to constrain τ_v since we do not have data in the region of the spectrum which it most affects. For non-starburst SEDs the

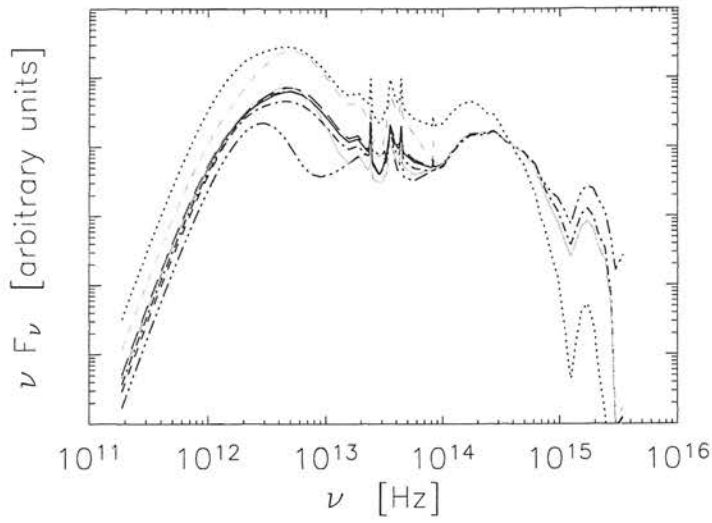

Figure 3.3: Examples of the affects of the model parameters on the resulting SEDs, see Section 3.3. Solid black line: $A_V = 1$, $f_{SB} = 1$, $\tau_v = 100$ (comparison model). Black dotted line: $A_V = 3$, $f_{SB} = 1$, $\tau_v = 100$. Dash-dot-dot-dot line: $t_m = 0$. Grey dashed line: $A_V = 1$, $f_{SB} = 5$, $\tau_v = 100$. Dashed line: $\chi = 10$. Dash-dot line: $\tau = 80$ Gyrs. Long dashed line: $\tau_{SB} = 80$ Myrs. Solid grey line: $\tau_v = 200$.

confidence region includes $A_V \pm 0.4$ and $\psi = 0 - 20$. We also note that since the effect of ψ on the shape of the model SEDs is not marked, it is difficult to constrain this parameter and the 1σ confidence region encompasses the whole allowed parameter range. However, since it has little effect on the shape of the model SED this will not affect the property estimates calculated from the models. Examples of these confidence regions for both the starburst (FN2 008) and non-starburst (FN2 001) galaxies are shown in Figures 3.4 to 3.7.

3.4 Spectral Energy Distributions

We have compared the observed spectral energy distributions with predictions from the radiative transfer models of Efstathiou and Rowan-Robinson (2003) and Efstathiou et al. (2000) described in Section 3.3.

Spectroscopic redshifts were used in the fitting of the model SEDs, where available from the ELAIS catalogue (Rowan-Robinson et al. 2004) or elsewhere. In other cases photometric redshifts were drawn from the ELAIS catalogue, and calculated bespoke

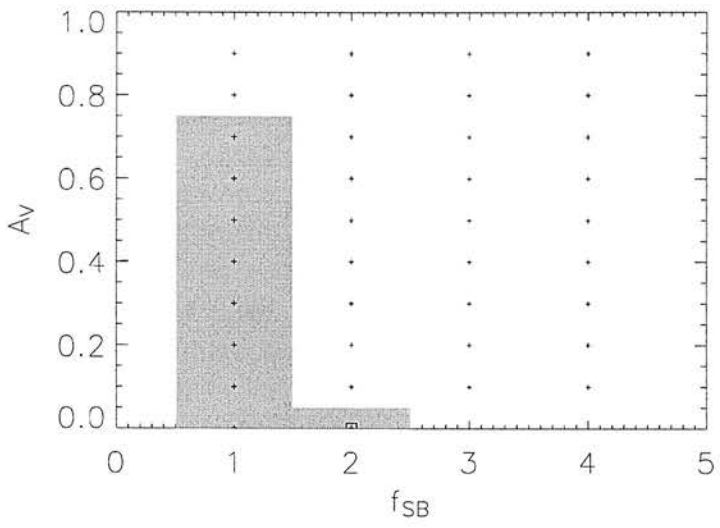

Figure 3.4: 1σ starburst model parameter confidence region for f_{SB} and A_V for FN2 008. Possible models parameter combinations are shown with small crosses (note $f_{SB} = 1 - 100$ and $A_V = 0 - 3$ were allowed), the best fit model is marked with a square and the 1σ confidence region is shaded.

by ImpZ (Babbedge et al. 2004) for the remainder. We note that since photometric redshifts are typically accurate to about 10 per cent in $(1 + z)$, for the spectroscopic redshifts we have available this implies an accuracy of around 0.11 to 0.14. Therefore, given that the solutions are based on a maximum of four optical bands, the agreement between our photometric and spectroscopic redshifts is reasonable; the largest margin of error being about 20 per cent in $(1+z)$ and it is also important to note that there are not catastrophic outliers which can occur due to degeneracies in the solution space of photometric codes. These small redshift errors do not affect the fitted SEDs significantly and the resulting errors on derived properties are much smaller than those incurred by the 1σ error on the fitted SED χ^2 value which are given for all derived quantities.

An aperture correction was applied to the sources' optical magnitudes to estimate their total integrated flux so that values were comparable to fluxes measured at other wavelengths. This was estimated using total SExtractor magnitudes in the r'–band for the optical objects from SExtractor run over the N2 INT–WFS r'–band image (provided by Eduardo González–Solares). The estimated total magnitudes are then

$$m_{tot} = m + (m_{r_{tot}} - m_r) \tag{3.4}$$

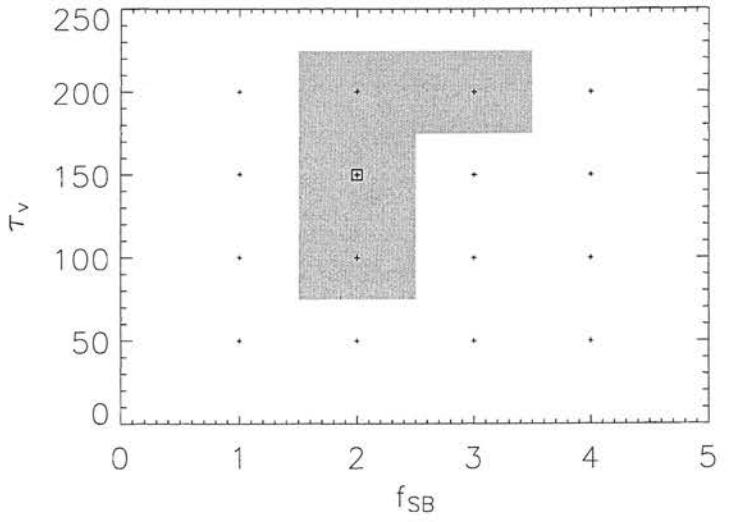

Figure 3.5: 1σ starburst model parameter confidence region for f_{SB} and τ_v for FN2 008. Possible models parameter combinations are shown with small crosses (note $f_{SB} = 1 - 100$ and $\tau_v = 50 - 200$ were allowed), the best fit model is marked with a square and the 1σ confidence region is shaded.

where m is the uncorrected optical magnitude in the optical band in question, m_r is the uncorrected r'–band magnitude and $m_{r_{tot}}$ is the total r'–band magnitude as estimated by SExtractor.

The N2 FIRBACK source SEDs are shown in Figures 3.8 and 3.9. They appear to fall into two population categories, the first being starbursting galaxies with their starburst region at a high optical depth (fitted model parameters shown in Table 3.3 and SEDs plotted in Figure 3.8) and the second dust obscured galaxies with quiescent star formation (fitted model parameters shown in Table 3.4 and SEDs plotted in Figure 3.9).

3.5 Star Formation Rates and Far-Infrared Luminosities

Star formation rates for the sources were calculated using both an optical U-band estimate (Cram et al. 1998) and an FIR estimate (Kennicutt 1998) given in equations 3.5 and 3.6.

$$\mathrm{SFR_U(M \geq 5\ M\odot)} = \frac{\mathrm{L_U}}{1.5 \times 10^{22}\mathrm{W\ Hz^{-1}}} 100\ \mathrm{M_\odot yr^{-1}} \qquad (3.5)$$

Table 3.3: Parameters for the starburst radiative transfer model SEDs fitted to the FIRBACK source SEDs.

FIRBACK	A_V	f_{SB}	τ_v
FN2 002	0.4	2.0	150
FN2 003	0.3	1.0	200
FN2 004	0.0	1.0	200
FN2 005	0.0	2.0	100
FN2 007	0.2	2.0	200
FN2 008	0.0	2.0	150
FN2 010	0.3	3.0	200
FN2 011	0.5	3.0	200
FN2 012	0.0	3.0	100
FN2 016	0.3	2.0	200
FN2 017	0.4	1.0	200
FN2 018	0.0	2.0	200
FN2 019	0.0	1.0	200
FN2 021	0.0	1.0	150
FN2 022	0.1	1.0	150
FN2 025	0.8	3.0	150
CFN2 029	1.1	10.0	200
CFN2 030	0.9	2.0	200
CFN2 034	1.0	1.0	50
CFN2 038	0.0	1.0	150
CFN2 039	0.0	1.0	100
CFN2 044	0.0	4.0	150
CFN2 047	0.0	1.0	50
CFN2 049	0.1	5.0	50

Table 3.4: Parameters for the quiescent radiative transfer model SEDs fitted to the FIRBACK source SEDs.

FIRBACK	A_V	ψ
FN2 000	10.0	20.0
FN2 001	0.4	17.0
FN2 015	1.1	20.0
FN2 020	3.7	0.5
FN2 023	0.2	0.5
CFN2 036	1.0	0.5

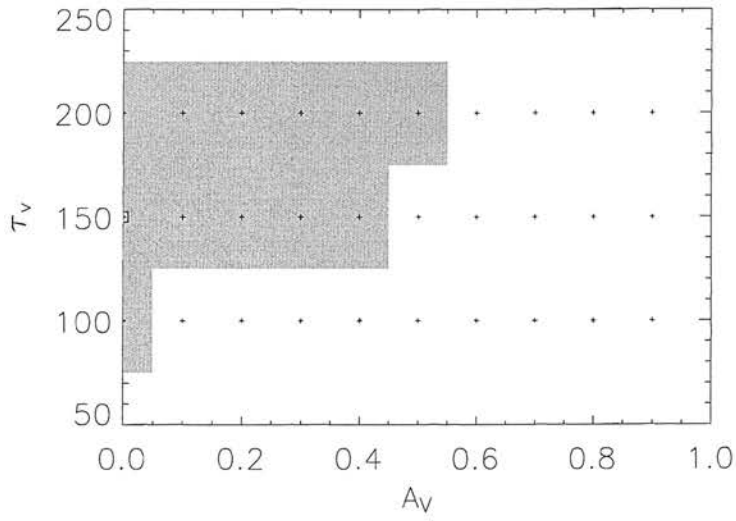

Figure 3.6: 1σ starburst model parameter confidence region for A_V and τ_v for FN2 008. Possible models parameter combinations are shown with small crosses (note $A_V = 0 - 3$ and $\tau_v = 50 - 200$ were allowed), the best fit model is marked with a square and the 1σ confidence region is shaded.

$$\mathrm{SFR_{FIR}} = 4.5 \times 10^{-44} \mathrm{L_{FIR}(ergs\ s^{-1})\ M_{\odot} yr^{-1}} \tag{3.6}$$

Where L_U is the U-band luminosity, estimated through interpolation of the SED models and L_{FIR} is the integrated infrared luminosity over 8-1000 μm. These both assume a Salpeter IMF. The chosen IMF affects the estimated star formation rates since we integrate over only a portion of the luminosity of the galaxy's constituent stars. The proportion of stars emitting with luminosities in the range in question will depend on the initial mass function, since it describes the initial distribution of stellar luminosities over the range, as discussed by Mann et al. (2002). The star formation rates for individual sources can be seen in Table 3.5. The distributions of star formation rates can be seen in Figure 3.10. There is a large spread in star formation rates with a number of galaxies with SEDs suggestive of a modest starburst. This implies that these galaxies are of low mass since their existing stellar population is dominated by the small starburst at infrared wavelengths. Using star formation rates estimated for SED models falling in the 1σ χ^2 range (see Section 3.3) the U–band and FIR SFRs are estimated to be accurate to 25 per cent and 50 per cent respectively. It can be seen in the lower panel of Figure 3.10 that for a substantial number of sources the FIR star formation rate

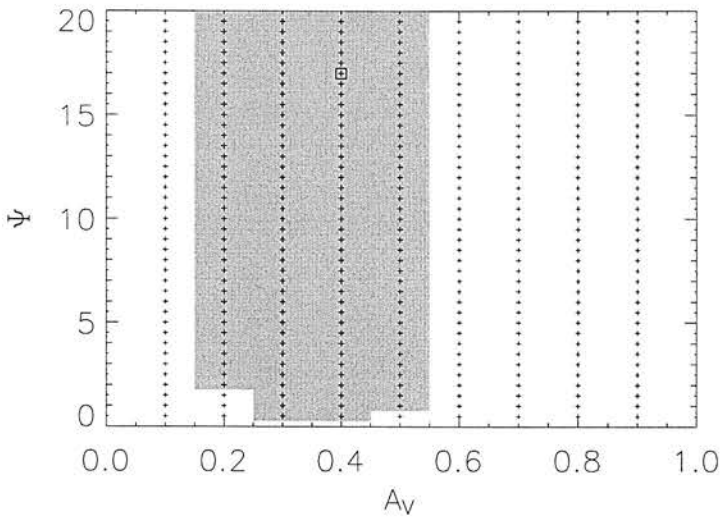

Figure 3.7: 1σ non–starburst model parameter confidence region for A_V and ψ for FN2 001. Possible models parameter combinations are shown with small crosses (note $A_V = 0 - 10$ and $\psi = 0.5 - 20.0$ were allowed), the best fit model is marked with a square and the 1σ confidence region is shaded.

estimates are significantly higher, by up to a factor of 4, than those using the U–band estimate. This is to be expected for such highly obscured objects.

The integrated far-infrared luminosities over the wavelength range 8-1000 μm were estimated. Results can also be seen in Table 3.5 and a distribution in Figure 3.11. These are estimated to be accurate to approximately 25 per cent.

3.6 Dust Temperatures and Masses

The dust in the galaxy radiative transfer models is at a range of temperatures, however, we computed an emission-weighted estimate via the fitting of a grey body to the far-infrared portion of the SED (from the peak of the SED to 1000 μm), see equation 3.7. This simply measures the colour temperature of the system and temperatures derived this way are unphysical simplifications. However, they do serve as a useful comparison with previous work and properties predicted using this method have been found to be consistent with those derived directly from complex radiative transfer models (Farrah

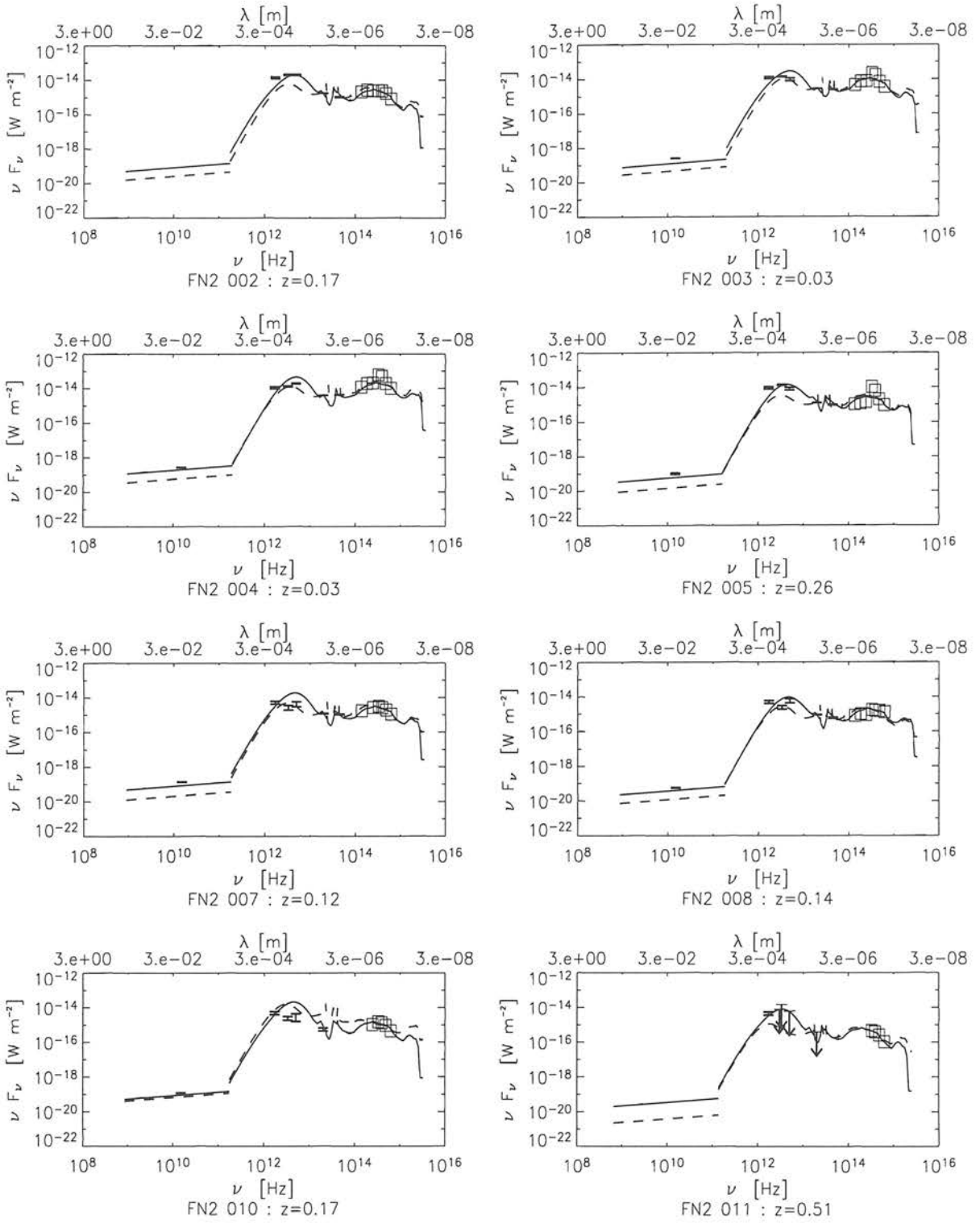

Figure 3.8: N2 FIRBACK sources fitted model SEDs. Best fit starburst models (solid line) and cirrus models (dashed line) are shown for each source.

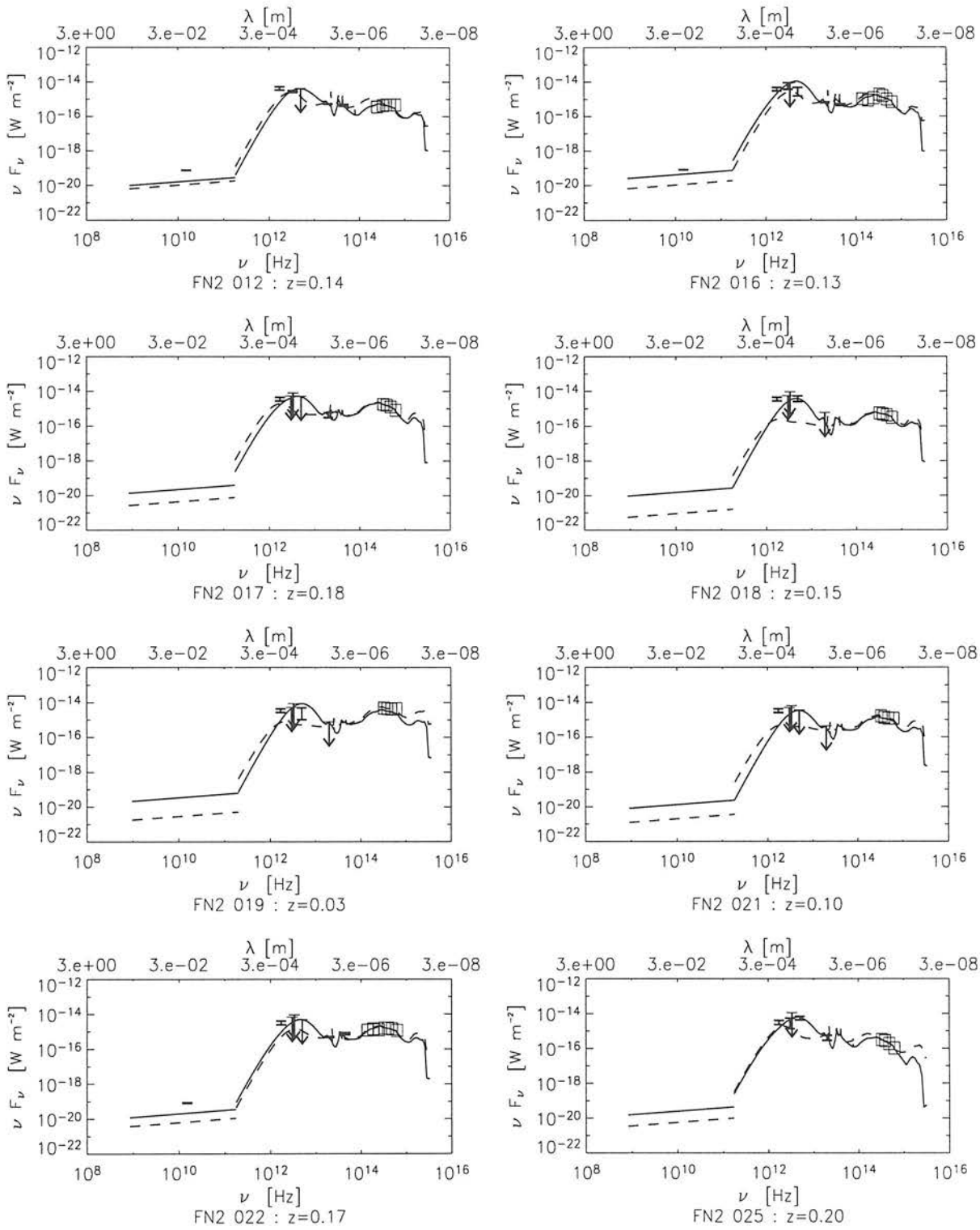

Figure 3.8: *continued*

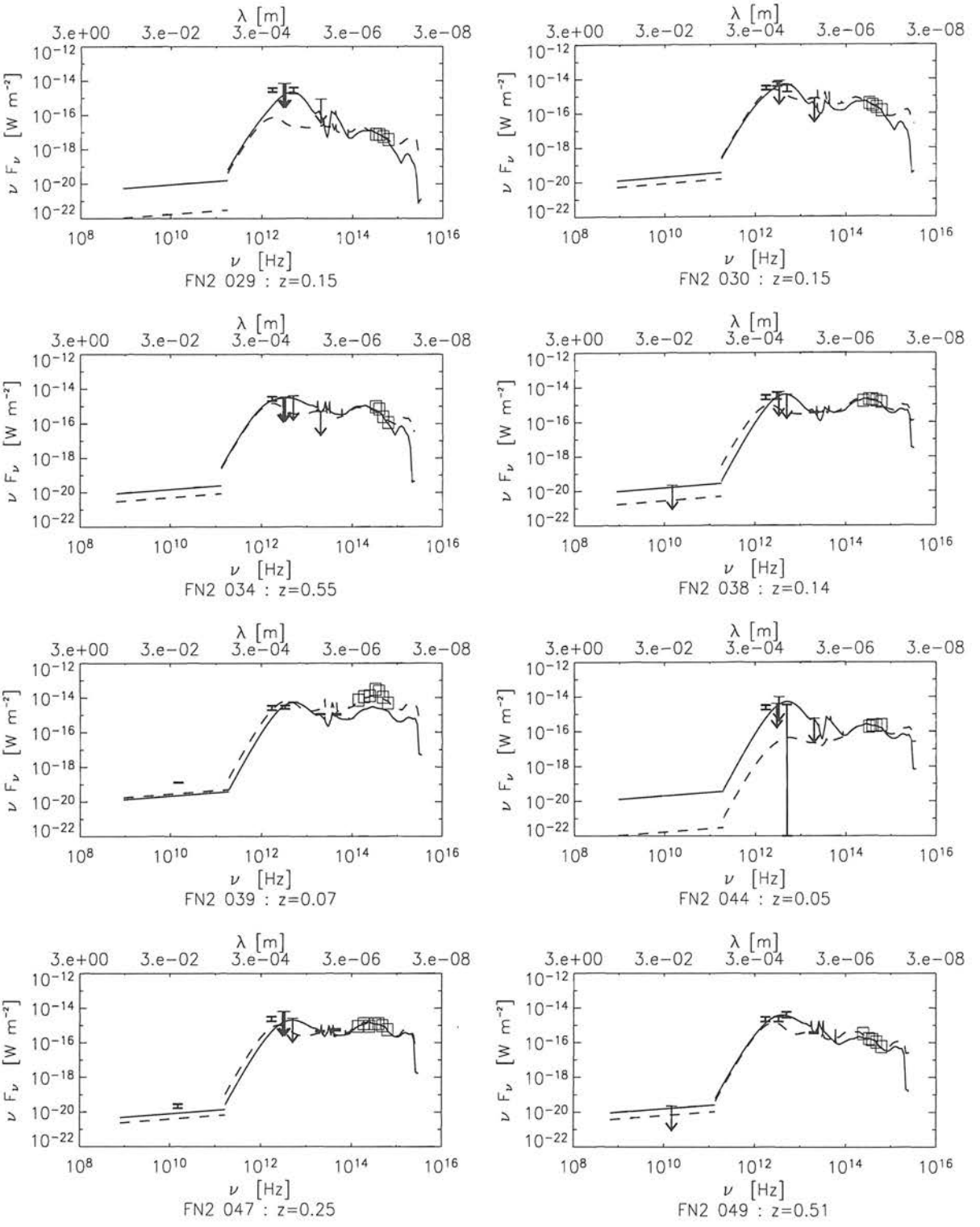

Figure 3.8: *continued*

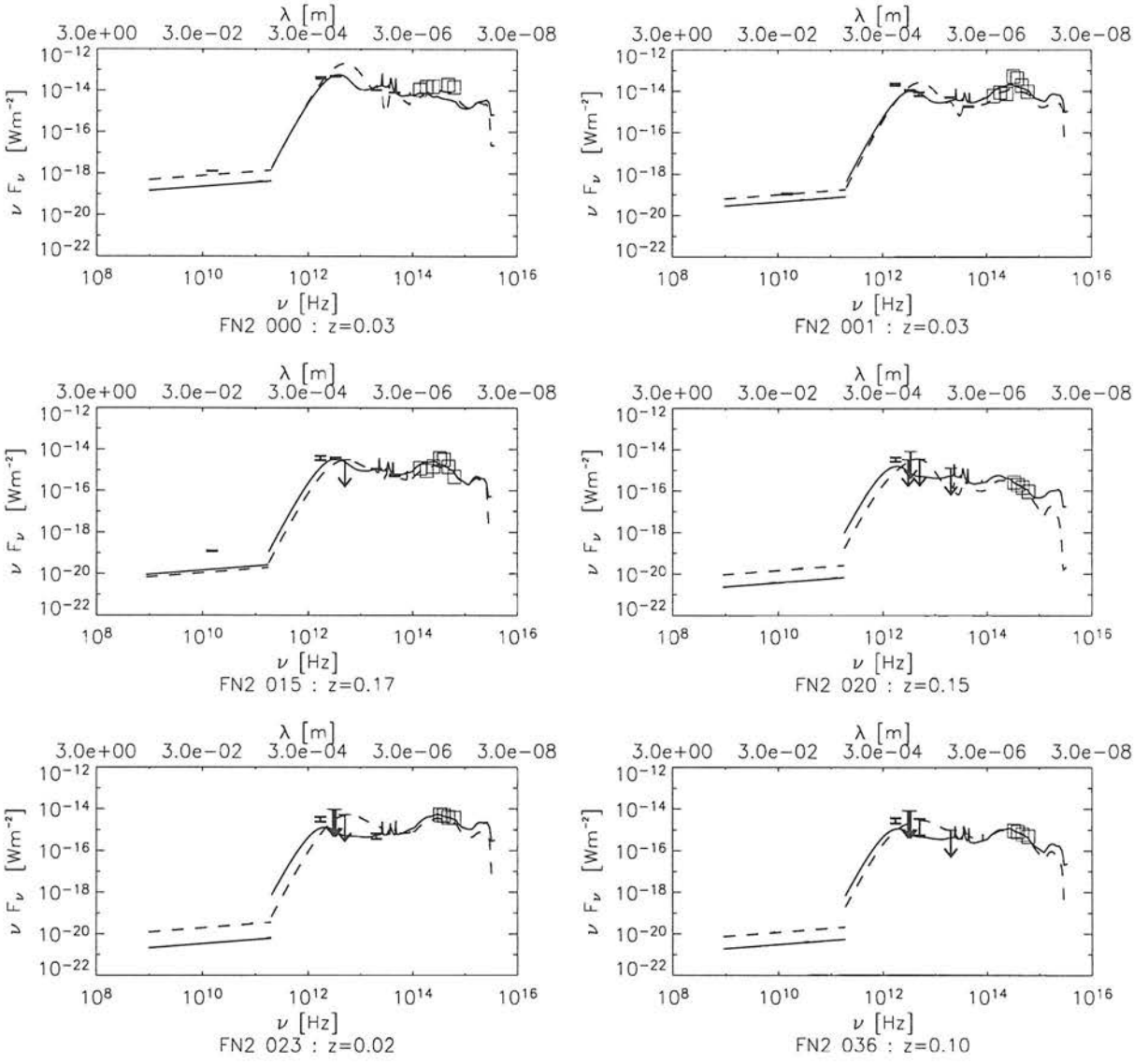

Figure 3.9: N2 FIRBACK sources fitted model SEDs. Cirrus models (solid line) and starburst models (dashed line) are shown.

Table 3.5: star formation rates and FIR luminosities of the starburst galaxies (first block) and cirrus galaxies (second block).

FIRBACK	u-band SFR	FIR SFR	L_{FIR}	z
	$M_\odot yr^{-1}$	$M_\odot yr^{-1}$	$L_\odot$	
FN2 002	36.9	93.1	4.87×10^{11}	0.17
FN2 003	3.8	3.6	1.83×10^{10}	0.03
FN2 004	9.4	5.4	2.85×10^{10}	0.03
FN2 005	131.4	169.1	9.18×10^{11}	0.26
FN2 007	22.3	39.5	2.07×10^{11}	0.12
FN2 008	24.4	27.3	1.46×10^{11}	0.14
FN2 010	21.7	88.9	4.68×10^{11}	0.17
FN2 011	104.8	391.9	2.04×10^{12}	0.51
FN2 012	6.3	13.1	7.07×10^{10}	0.14
FN2 016	12.6	25.6	1.33×10^{11}	0.13
FN2 017	32.5	28.1	1.41×10^{11}	0.18
FN2 018	8.4	12.2	6.53×10^{10}	0.15
FN2 019	1.7	1.0	5.22×10^{9}	0.03
FN2 021	7.9	4.8	2.59×10^{10}	0.10
FN2 022	32.6	22.3	1.18×10^{11}	0.17
FN2 025	5.8	39.7	2.07×10^{11}	0.20
CFN2 029	0.1	7.2	3.79×10^{10}	0.15
CFN2 030	3.2	16.8	8.60×10^{10}	0.15
CFN2 034	138.2	296.6	1.48×10^{12}	0.55
CFN2 038	25.6	11.6	6.19×10^{10}	0.14
CFN2 039	7.7	4.1	2.23×10^{10}	0.07
CFN2 044	0.4	1.8	9.41×10^{9}	0.05
CFN2 047	63.2	27.2	1.49×10^{11}	0.25
CFN2 049	49.1	252.8	1.39×10^{12}	0.51
FN2 000	2.8	7.9	3.97×10^{10}	0.03
FN2 001	9.3	1.7	8.37×10^{9}	0.03
FN2 015	29.1	18.8	9.58×10^{10}	0.17
FN2 020	3.0	7.5	2.56×10^{10}	0.15
FN2 023	1.1	0.1	3.87×10^{8}	0.02
CFN2 036	4.3	2.5	8.82×10^{9}	0.10

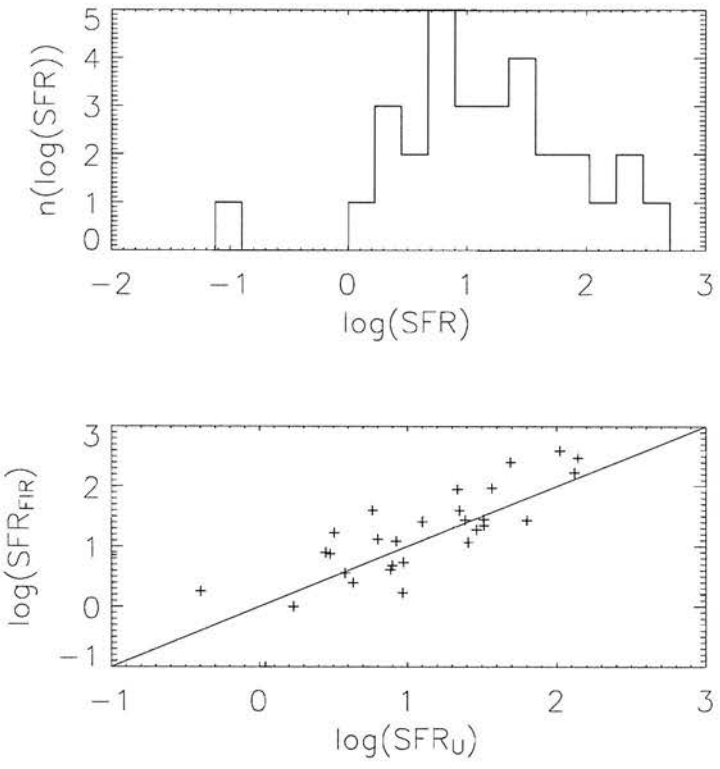

Figure 3.10: Top window:Far-infrared log(star formation rate) distribution ($M_\odot yr^{-1}$). Bottom window: Far-infrared log(star formation rate) estimates vs. U–band log(star formation rate) estimates.

et al. 2003).

$$F_\nu \propto \frac{\nu^{\beta+3}}{e^{\frac{h\nu}{kT}} - 1} \tag{3.7}$$

T is the dust temperature which controls the position of the maximum of the long wavelength SED, β is the frequency dependence of the dust grain emissivity and controls the gradient of the long wavelength slope of the grey body curve. Fitting over this wavelength range allows us to constrain both the temperature and β, without reaching the far-infrared where the dust may go optically thick and no longer be well described by a greybody model. The derived values for the sources can be seen in Table 3.6. Using the method described above (Section 3.5) for the starburst models these are estimated to be accurate to $T \pm 3$ K and $\beta \pm 0.5$, in the non-starburst case these estimates are accurate to $T \pm 14$ K and $\beta \pm 0.1$.

The distribution of the dust temperatures for the N2 FIRBACK source galaxies is shown in Figure 3.12. The starburst galaxy dust temperatures are slightly higher than

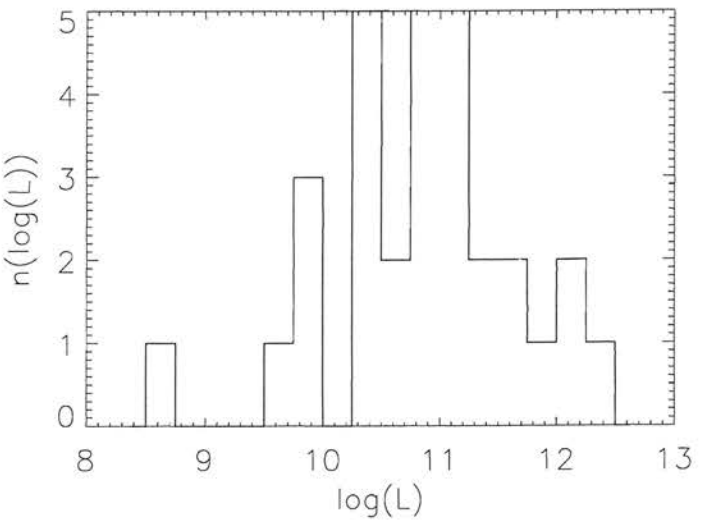

Figure 3.11: Distribution of log(far-infrared luminosities ($L_\odot$)).

those found by Sajina et al. (2003) ($\sim 20-30$ K) for a sample of N1 FIRBACK sources but have similar emissivity coefficients: Sajina et al. (2003) found $\beta \sim 1.5-1.7$.

Dust masses were estimated using the submillimetre prescription (Hildebrand 1983) at a rest wavelength of 450 μm:

$$M_{dust} = \frac{S_{\nu_0} D_L^2}{(1+z)\kappa(\nu_r) B(\nu_r, T_{dust})} \tag{3.8}$$

$$\kappa(\nu_r) = 0.067 \left(\frac{\nu_r}{2.5 \times 10^{11}} \right)^{\beta} \tag{3.9}$$

where ν_0 and ν_r are the observed and rest frame frequencies respectively, S_{ν_0} is the observed flux at ν_0, estimated through interpolation of the SED models, and $B(\nu_r, T_{dust})$ is the Planck function in the rest frame. T_{dust} and β are the dust temperature and dust grain emissivity estimated via the greybody fitting. $\kappa(\nu_r)$ is known as the mass absorption coefficient and has units of $m^2 kg^{-1}$. This approach assumes that the galaxy is optically thin at submillimetre wavelengths. Resulting dust masses, calculated at a wavelength of 450 μm (accurate to approximately 25 per cent), are shown in Table 3.6. We note that the dust masses follow the trend of the far-infrared luminosities and star formation rates, those with high SFRs and L_{FIR} having larger dust masses.

Table 3.6: Greybody parameters and dust masses for the starburst galaxies (first block) and cirrus galaxies (second block).

FIRBACK	T (K)	β	M_{dust}($M_\odot$)
FN2 002	40	1.4	8.0×10^7
FN2 003	36	1.5	4.3×10^6
FN2 004	41	1.9	1.4×10^6
FN2 005	40	1.9	2.9×10^7
FN2 007	42	1.5	2.5×10^7
FN2 008	39	1.9	6.5×10^6
FN2 010	45	1.4	5.4×10^7
FN2 011	41	1.4	2.5×10^8
FN2 012	39	1.9	3.0×10^6
FN2 016	42	1.4	2.1×10^7
FN2 017	36	1.4	3.9×10^7
FN2 018	39	2.0	2.4×10^6
FN2 019	41	1.9	2.5×10^5
FN2 021	41	1.9	9.9×10^5
FN2 022	41	1.6	1.0×10^7
FN2 025	40	1.3	4.5×10^7
CFN2 029	49	1.2	5.6×10^6
CFN2 030	38	1.3	2.5×10^7
CFN2 034	35	1.1	6.5×10^8
CFN2 038	40	1.9	2.6×10^6
CFN2 039	40	1.9	9.2×10^5
CFN2 044	41	1.9	4.2×10^5
CFN2 047	41	1.7	6.6×10^6
CFN2 049	44	1.5	6.6×10^7
FN2 000	31	1.9	7.2×10^6
FN2 001	30	1.9	1.7×10^6
FN2 015	31	1.9	1.4×10^7
FN2 020	17	1.8	1.6×10^8
FN2 023	16	2.0	2.0×10^6
CFN2 036	17	1.8	5.1×10^7

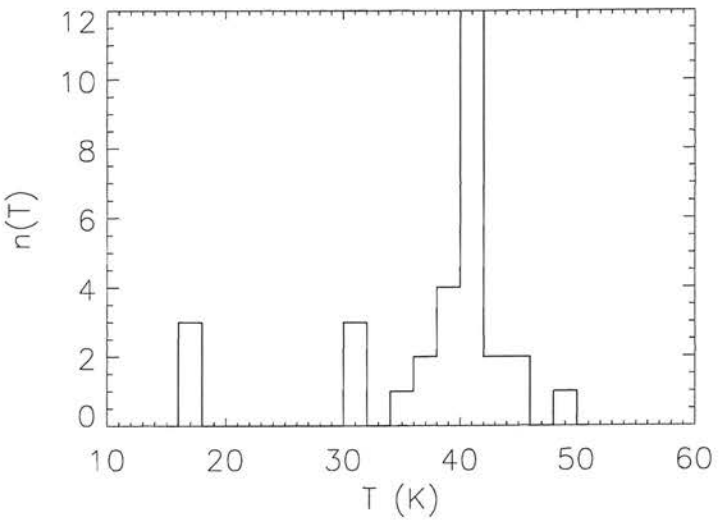

Figure 3.12: Distribution of dust temperatures.

3.7 The Nature of the N2 FIRBACK Sources

We have found bright galaxy optical identifications for 31 out of the 55 N2 ISO 175 μm sources and stellar identifications for 2 sources. 17 of those with galaxy optical IDs were associated with sources from the band-merged ELAIS catalogue. We then compared the spectral energy distributions of the sources with predictions from the radiative transfer models of Efstathiou and Rowan-Robinson (2003) and Efstathiou et al. (2000).

22 of the 175 μm sources have not been confidently associated with optical counterparts and therefore their nature is still unknown. However, only 4 of these sources are in the main FIRBACK catalogue, the remaining 18 are listed in the supplementary catalogue which contains detections at lower signal–to–noise and therefore these could just be spurious detections. However, these could also be due to sources having multiple IDs or their being associated with heavily extincted objects which were not in the optical catalogue.

24 galaxies have been classed as starbursting. They have a range of far-infrared luminosities from 10^9 $L_\odot$ to 10^{12} $L_\odot$ corresponding to a range of star formation rates from $\sim$ 0.1 to $\sim$ 400 $M_\odot yr^{-1}$. Although those at the lower end of this range are not forming stars at a dramatic rate their small starburst component is dominating over the existing population at far-infrared wavelengths. This implies that the 175 μm

FIRBACK sample contains a significant number of low mass galaxies. The majority of these galaxies are spirals and a number are S0 galaxies.

Sources FN2 005, 011, CFN2 034 and 049 have $L_{FIR} \sim 10^{12}$ $L_{\odot}$ and high star formation rates, we therefore suggest that these are in the low redshift tail of the rapidly evolving ULIRG population. This is in agreement with the findings of Sajina et al. (2003) who found one sixth of their 175 μm sample to fall into this category. We also suggest that those galaxies with moderate star formation rates and far-infrared luminosities of $\sim 10^{11}$ $L_{\odot}$ at moderate z (≥ 0.12) are LIRGs (FN2 002, 007, 008, 010, 016, 017, 022, 025 and CFN2 047). Those with $z \leq 0.12$ are more normal galaxies perhaps with lower masses (FN2 003, 004, 012, 018, 019, 021, CFN2 029, 030, 038, 039, 044).

The estimated dust temperatures of the starburst galaxies are 30-49 K which is relatively low compared to that of standard ULIRGs, which have $T_D \sim 50$ K. The most luminous of these galaxies also have large dust masses adding weight to the suggestion that these cold, dusty, luminous objects have large masses of dust heated at a moderate intensity (Chapman et al. 2002).

A further 6 sources have non-starburst SEDs. These sources have far-infrared luminosities of $10^8 - 10^{10}$ $L_{\odot}$ and star formation rates of < 30 $M_{\odot}yr^{-1}$. We suggest that these are quiescently starforming galaxies (FN2 000, 001, 015, 020, 023, 036). These galaxies have low dust temperatures of 16-31 K and dust masses of $10^6 - 10^8$ $M_{\odot}$. 3 of these galaxies are spirals, 1 is an interacting galaxy pair, 1 is either an irregular or a spiral with tidal tail and the morphology of the remaining source was unable to be deduced.

Lagache et al. (2003) predicted 62 per cent of 170 μm sources would have $z < 0.25$ and the remainder $0.8 < z < 1.2$. We find only one source with $S > 180$ mJy with $z > 0.25$ and therefore have $\gg 62$ per cent with $z < 0.25$. However, our mix of higher redshift ULIRGs and lower redshift starburst galaxies is in good qualitative agreement with the predictions of the Lagache et al. (2003) starburst templates as adopted by Sajina et al. (2003). Lagache et al. (2003) also state that to gain agreement between their models and observations they require a cold local population which is what has been found.

There has also been recent investigation into the submillimetre properties of the local universe through the SCUBA Local Universe Galaxy Survey (Dunne and Eales 2001; Dunne et al. 2000). This survey found a population of galaxies with $L_{FIR} = 10^{10} - 10^{11}$ $L_{\odot}$ and dust masses of $10^7 - 10^8$ $M_{\odot}$. The population was fitted with a

two-component dust model to discover cold dust temperatures of ~ 21 K and hot dust temperatures of $30-50$ K. It is clear that the local 175 μm FIRBACK population has very similar properties to these galaxies. The sources of the far-infrared background are therefore offering a stepping stone between high redshift SCUBA galaxies and the local universe.

3.8 Summary of the Nature of the N1 FIRBACK Sources

The FIRBACK sources in the N1 ELAIS region were examined by Dennefeld et al. (2005) (a full version of this paper is given in Appendix B) independently but in parallel with our work on the N2 sources. We provided them with multicolour optical associations with the INT–WFS catalogue using the Mann et al. (1997, 2002) likelihood ratio technique. These were then associated with ELAIS 15 μm sources, SCUBA 450 μm and 850 μm detections, IRAS 60 μm and 100 μm detections and radio data from the VLA by Ciliegi et al. (1999) and the FIRST survey (Becker et al. 1995).

The investigation was concentrated on the main N1 FIRBACK catalogue ($S >$ 180mJy) which contained 56 sources and were able to make firm identifications for 28 of these. Optical spectra for these sources were obtained from the Haute–Provence Observatory (CNRS, France) and a very similar redshift distribution to that for the N2 sources was found with approximately 80 per cent of the galaxies having $z <$ 0.25. Dennefeld et al. (2005) also compared the multiwavelength data with template SEDs from Lagache et al. (2003) in order to estimate far-infrared luminosities and star formation rates.

The majority of the FIRBACK N1 sources were found to be quiescently star forming galaxies whose spectra were fit by a cold galaxy template. These sources had far–infrared colours of $F_{170\mu\mathrm{m}}/F100\mu\mathrm{m} >\sim 0.7$ and $F_{100\mu\mathrm{m}}/F60\mu\mathrm{m} >\sim 1.5$ further illustrating that these are cold galaxies dominated by a cold dust component at $T \sim 20-40$ K. The brightest FIRBACK N1 sources were LIRGs with moderate SFRs of 5-10 $\mathrm{M}_{\odot}\mathrm{yr}^{-1}$. The ratio of total far–infrared luminosity to K-band luminosity was used to indicate the ratio of present–to–past star formation, this was ~20-40 for the N1 sources showing that their energy output is dominated by the present star formation. Five of the galaxies were more active than average, one of these with a large far–infrared luminosity bringing it into the ULIRG class.

It was concluded that the bright FIRBACK N1 sources were fainter versions of the cold galaxies found in the ISOPHOT 170 μm serendipity survey (Stickel et al. 2000).

3.8. SUMMARY OF THE NATURE OF THE N1 FIRBACK SOURCES

The emission lines, reddening and star formation rates found in the optical spectra suggested that the N1 sources resemble faint IRAS galaxies found by Bertin et al. (1997) and Patris et al. (2003). Dennefeld et al. (2005) noted that there is, therefore, a large population of local cold galaxies with moderate star formation rates whose contribution to the global star formation rate has been previously overlooked. This population must be taken into account in evolutionary models for infrared galaxies as it will affect the z–distribution of 170 μm sources predicted. Models considering only starburst galaxies will predict 4σ FIRBACK galaxies with a z–distribution biased to much higher redshifts than observed.

The aim of the FIRBACK survey was to resolve the cosmic far–infrared background into discrete sources and hence discover its origin, however, the bright N1 sources were found to have mid-to-far–infrared colours that are different from those of the cosmic infrared background and therefore the bulk of contributors to the cosmic infrared background have not yet been found. They expect these to be more distant, active galaxies.

It is clear that our conclusions about the nature of the N2 FIRBACK population and its implications for evolutionary models are consistent with those made for the N1 population.

It is more important to have beauty in one's equations than to have them fit experiment.

Paul Dirac 1902-84

4

Machine Learning and Associations

4.1 Introduction

As described in Chapter 2 the magnitude distribution of the true source counterparts, $q(f)$, may be unknown for some source catalogues. This is a common situation, since it is often the case that we wish to study a previously unknown population e.g. through observations with a new instrument. In that case we have no prior knowledge about the properties of the source counterparts we are searching for in a catalogue at a different wavelength. As described in Chapter 2, it was necessary to make some assumption about the magnitude distribution, $q(f)$, in the likelihood ratio source association method. We therefore wanted to try a totally new method in which this assumption would not need to be made and instead the magnitude distribution could be 'learned' from the data.

Through consultation with colleagues in the University of Edinburgh's School of Informatics we decided that the Expectation Maximisation (EM) algorithm from the machine learning community might be appropriate and developed a new astronomical application in collaboration with Amos Storkey. This is described in Section 4.2.2, and in Storkey et al. (2005).

The EM algorithm was first tested on finding optical associations for the ISO FIRBACK 175 μm sources in the INT–WFS catalogue (see Chapter 2 for a description of these data sets). Since we had already found IDs for these sources using the likelihood ratio method (see Chapter 2) we were able to compare the effectiveness of these two different methods and the IDs made. The application of our new algorithm to the FIRBACK data is described in Section 4.2 and a comparison with the likelihood ratio method in Section 4.3.

A group at the University of Queensland have studied (Rohde et al. 2005) the use of other machine learning techniques for finding associations for the HI Parkes All–Sky Survey (HIPASS) catalogue. To facilitate a direct comparison of the various association methods we applied both the likelihood ratio and EM methods to their data set. HIPASS is described in Section 4.4.1 and the SuperCOSMOS optical data with which associations were sought is described in Section 4.4.2, while Section 4.4.4 describes the Rohde et al. (2005) machine learning association project. The other advantage of this data set was that part of the HIPASS catalogue (HICAT) had already been identified with optical counterparts through analysis of radial velocity data to produce HOPCAT (see Section 4.4.1, Doyle et al. 2005). This, therefore, gave another means of testing if the likelihood ratio and the EM algorithm were giving the 'correct answers'. The application of the EM algorithm to the HIPASS survey is described in Section 4.4 and a comparison of both these methods and the Rohde et al. (2005) Support Vector Machine (SVM) in Section 4.5.

Both the FIRBACK and HIPASS datasets pose unusually challenging association problems due to their large positional errors, leading to numerous counterpart candidates to choose between within their error ellipses. These therefore exemplify the kind of situation where an automated statistical approach to finding IDs must be followed.

4.2 FIRBACK Associations with the EM Algorithm

The EM (Expectation–Maximisation) algorithm is a standard method, used in the fields of statistics and informatics, for finding maximum likelihood estimates when there are incomplete or missing data, but which has never been utilised in the matching of astronomical data before. Its development is usually credited to Dempster et al. (1977) but in fact it was first independently developed by the astronomer L. B. Lucy in 1974 for the 'rectification of observed distributions' (Lucy 1974). He suggested that it could be used for beam smoothing in radio–astronomy, however its value appears to have

been overlooked and it didn't catch on at the time. We have used it in an entirely new way to learn the unknown magnitude distribution of the 175 μm source optical counterparts and hence enable us to select IDs for the sources.

4.2.1 Probability Distribution in the Data

Before describing the EM algorithm, with help from Amos Storkey we set up the problem we will use it to solve. We have a number of sources, labelled r, for which we seek optical associations within a search radius R_r. Optical objects within R_r are labelled i, each having an associated displacement relative to the source, x_i^r, and a magnitude m_i^r. An indicator function, z_i^r, is used to label whether an object is the true ID ($z_i^r = 1$) or it is a background object ($z_i^r = 0$). We also allow for the possibility that there is no optical counterpart for the source, in which case an additional indicator function, z_0^r, is unity. Suppose there are N_r objects within the search radius R_r, then we require $\sum_{i=0}^{N_r} z_i^r = 1$; this allows at most one optical counterpart per source.

The probability, $P(x_i^r|z_i^r = 1)$, that an optical object, i, has a position x_i^r given that it is associated with the source r, is assumed to have a Gaussian distribution, the covariance of which is the error on the position of the source, σ. $P(x_i^r|z_i^r = 0)$, the probability of an optical object i having a position x_i^r given that it is not associated with the source (i.e. it is just a background object) is estimated as $\mu_r = 1/(\pi R_r^2)$

The magnitude distribution, $P(m_i^r|z_i^r = 0)$, for the background objects is known; it is just the normalised overall magnitude distribution of the optical catalogue. However, the magnitude distribution of the optical counterparts for the sources, $P(m_i^r|z_i^r = 1)$, is that which is unknown and must be 'learned'. We assume it may be represented by a histogram with B bins of width w:

$$P(m_i^r|z_i^r = 1) = \sum_{t=1}^{B} h_t T_t(m_i^r) \tag{4.1}$$

T_t is a top hat function which is 1 if m_i^r lies within the ith bin and 0 otherwise. We set $w \sum_t h_t = 1$ such that h_t is the normalised bin height.

The probability that a source has no optical counterpart is also unknown and set to γ: i.e. $P(z_0^r = 1) = \gamma$. We then assume that if there is an optical counterpart, each of the candidates has an equal probability of being the true ID. Therefore, $P(z_i^r = 1) = \frac{(1-\gamma)}{N_r}$.

4.2.2 Setting up the EM Algorithm

Having set up the problem, we now describe the EM algorithm, the iterative method used to learn $P(m_i^r|z_i^r = 1)$. This was done with help from Amos Storkey.

The EM algorithm utilises the KL (Kullbeck-Leibler) divergence which measures the 'distance' between two probability densities. This quantity, also known as the relative entropy, is given by

$$KL(W(x_i)||Y(x_i)) = \sum_i W(x_i) \ln \frac{W(x_i)}{Y(x_i)} \tag{4.2}$$

i.e. it is the expectation with respect to $W(x)$ of the difference between $\ln W(x)$ and $\ln Y(x)$. Note that the KL divergence cannot take a negative value and is at its minimum value of 0 when $W = Y$.

Consider a situation in which we have some data, D, and from these wish to deduce the best values for certain parameters, Θ, in a model describing the data. To do this we wish to maximise the likelihood $P(D|\Theta)$, the probability of the data given the parameters. To set this up we take the KL divergence of the posterior distribution, $P(Z|D,\Theta)$, where Z can be any property of the sources in question, and some arbitrary distribution $Q(Z)$.

$$\begin{aligned} KL(Q(Z)||P(Z|D,\Theta)) &= \langle \ln \frac{Q(Z)}{P(Z|D,\Theta)} \rangle_{Q(Z)} & (4.3) \\ &= \langle \left(\ln Q(Z) - \ln \left(\frac{P(Z,D|\Theta)}{P(D|\Theta)} \right) \right) \rangle_{Q(Z)} & (4.4) \end{aligned}$$

since $P(X,Y|T) = P(X|Y,T)P(Y|T)$.

$$\begin{aligned} KL(Q(Z)||P(Z|D,\Theta)) &= \langle (\ln Q(Z) - \ln P(Z,D|\Theta) + \ln P(D|\Theta)) \rangle_{Q(Z)} & (4.5) \\ &= \langle (\ln Q(Z) - \ln P(Z,D|\Theta)) \rangle_{Q(Z)} + \ln P(D|\Theta) & (4.6) \end{aligned}$$

since the likelihood $P(D|\Theta)$ is independent of Z, so

$$\ln P(D|\Theta) = \langle (\ln P(Z,D|\Theta) - \ln Q(Z)) \rangle_{Q(Z)} + KL(Q(Z)||P(Z|D,\Theta)) \tag{4.7}$$

We can now use the EM algorithm to maximise this log likelihood via finding the best values for the parameters, Θ.

Since $Q(Z)$ is independent of Θ we set $Q(Z) = P(Z|D,\Theta)$, the posterior distribution, hence reducing the KL divergence to its minimum value, 0. We then alter the parameters to maximise $\langle (\ln P(Z,D|\Theta) \rangle_{Q(Z)}$ and hence maximise the log–likelihood, $\ln P(D|\Theta)$. Having set the KL divergence to zero ensures that when the parameters

are adjusted we do not inadvertently reduce the KL divergence and consequently the log–likelihood, since the divergence cannot decrease below 0. Since Q does not depend on the parameters, Θ, it can be reset to the new value of the posterior given the new parameters without affecting the log–likelihood. This therefore gives an iterative procedure, summarised below, for maximising the log–likelihood by finding the best values for the parameters, Θ.

1. Initialise the parameters, Θ.

2. Set Q to the current estimate of the posterior distribution for Z, i.e. $Q(Z) = P(Z|D,\Theta)$.

3. Maximise $\langle(\ln P(D,Z|\Theta))\rangle_Q(z)$ given this value of $Q(Z)$ by altering the values of Θ.

4. Repeat steps 2 and 3 until suitably converged.

We now need to set up the EM algorithm in terms of the probability distributions in our data described in Section 4.2.1.

Since each source is independent,

$$\begin{aligned}\langle(\ln P(D,Z|\Theta))\rangle_{Q(Z)} &= \langle(\ln \prod_r P(D^r,Z^r|\Theta))\rangle_{Q(Z)} \\ &= \sum_r \langle(\ln P(D^r,Z^r|\Theta))\rangle_{Q(Z)} \end{aligned} \tag{4.8}$$

where Z^r is the set of all z_i^r for a given r and D^r is the same for the data, so

$$P(D^r,Z^r|\Theta) = P(Z^r)\prod_{i=1}^{N_r} P(x_i^r|z_i^r)P(m_i^r|z_i^r) \tag{4.9}$$

If we chose a Z^r from $P(Z^r)$ this equation will connect either 1 or 0 optical objects with that source. Then for all the optical objects their position and magnitude are sampled from the background distributions, if their $z_i^r = 0$, or from the counterpart distributions, if their $z_i^r = 1$. If $z_0^r = 1$, i.e. there is no optical counterpart, then x and m are sampled from the background distribution. We can substitute in $P(z_i^r = 1) = \alpha_i$, where, as described earlier, $\alpha_i = \gamma$ for $i = 0$ and $\alpha_i = (1-\gamma)/N_r$ for $i > 0$, to obtain

$$\begin{aligned}P(D^r,Z^r|\Theta) = \left(\prod_{i=0}^{N_r}\alpha_i^{z_i^r}\right)\prod_{i=1}^{N_r}(& P(x_i^r|z_i^r=1)^{z_i^r}P(m_i^r|z_i^r=1)^{z_i^r} \\ & P(x_i^r|z_i^r=0)^{(1-z_i^r)}P(m_i^r|z_i^r=0)^{(1-z_i^r)})\end{aligned} \tag{4.10}$$

The indicator variables z_i^r are used to pull out the correct components i.e. distributions for the true counterparts are raised to the power z_i^r such that they only contribute when

the ith object is the ID for that source and similarly background distributions are raised to the power $(1 - z_i^r)$.

We now substitute into this equation the forms of the distributions that we set out in Section 4.2.1:

$$P(D^r, Z^r|\Theta) = \left(\prod_{i=0}^{N_r} \alpha_i^{z_i^r}\right) \prod_{i=1}^{N_r} (P(x_i^r|z_i^r = 1)^{z_i^r} \left(\sum_t h_t T_t(m_i^r)\right)^{z_i^r} \mu_r^{(1-z_i^r)} P(m_i^r|z_i^r = 0)^{(1-z_i^r)}) \quad (4.11)$$

Taking logs and substituting into equation 4.8 gives:

$$\begin{aligned} \langle \ln P(D^r, Z^r|\Theta)\rangle_{Q(Z)} = \sum_r (\sum_{i=0}^{N_r} < z_i^r > \ln\alpha_i + \sum_{i=1}^{N_r} (< z_i^r > \ln P(x_i^r|z_i^r = 1) + \\ + < z_i^r > \ln \left(\sum_t h_t T_t(m_i^r)\right) + (1 - < z_i^r >)\ln\mu_r + \\ + (1 - < z_i^r >)\ln P(m_i^r|z_i^r = 0))) \end{aligned} \quad (4.12)$$

The expectation of an indicator variable is its probability of being 1. Therefore,

$$< z_i^r >= P(z_i^r = 1|D^r, \Theta) \quad (4.13)$$

and since $P(z_i^r|D^r, \Theta) \propto P(z_i^r = 1, D^r|\Theta) = P(Z^r, D^r|\Theta)$, we can calculate the expectation $< z_i^r = 1 >$ from equation 4.11. $P(Z^r, D^r|\Theta)$ is the probability of a particular Z^r and the data D^r being generated from the source for i and the background for the rest. Therefore, this is just a less general form of equation 4.11 i.e. since we know which $z_i^r = 1$ we can remove the components raised to z_i^r where $z_i^r = 0$. We can calculate this for every i and obtain the constant of proportionality since probabilities must sum to 1. For $i > 0$ we have:

$$\begin{aligned} < z_i^r = 1 > &= P(z_i^r = 1|D^r, \Theta) \\ &\propto P(z_i^r = 1, D^r|\Theta) \\ &\propto \alpha_i P(x_i^r|z_i^r = 1) \left(\sum_t h_t T_t(m_i^r)\right) \prod_{j=1, j\neq i}^{N_r} (\mu_r P(m_j^r|z_j^r = 0)) \end{aligned} \quad (4.14)$$

Similarly for $< z_0^r = 1 >$ i.e. $i = 0$:

$$\begin{aligned} < z_0^r = 1 > &= P(z_0^r = 1|D^r, \Theta) \\ &\propto P(z_0^r = 1, D^r|\Theta) \\ &\propto \alpha_0 \prod_{i=1}^{N_r} (\mu_r P(m_i^r|z_i^r = 0)) \end{aligned} \quad (4.15)$$

To perform the EM algorithm we wish to maximise $\langle \ln P(D^r, Z^r|\Theta)\rangle_{Q(Z)}$ by altering the parameters. We therefore take the derivative with respect to h_t, adding a Lagrange multiplier term, $\lambda(w\sum h_t - 1)$, to account for the normalisation constraint on h_t.

$$\frac{\partial}{\partial h_t}\langle \ln P(D^r, Z^r|\Theta)\rangle_{Q(Z)} = \sum_r \sum_{i=1}^{N_r} < z_i^r > \frac{T_t(m_i)}{h_t} + \lambda w \tag{4.16}$$

Setting this to zero and solving for h_t gives:

$$h_t \propto \sum_r \sum_{i=1}^{N_r} < z_i^r > T_t(m_i) \tag{4.17}$$

where the constant of proportionality can be found using the normalisation constraint.

This is repeated to find the parameter γ:

$$\gamma = \frac{\sum_r < z_0^r >}{\sum_r (< z_0^r > + \sum_{i=1}^{N_r} < z_i^r >)} \tag{4.18}$$

We note that this is effectively the average value of $< z_0^r >$ since the value of the denominator is just N_r.

Final Procedure

We now have all the information needed to implement the EM algorithm. In analogy to the summary given earlier, the application of the algorithm for our data is summarised below:

1. Initialise the parameters h_t and γ.

2. Calculate the marginal posteriors $P(z_i^r = 1|D^r, \Theta) =< z_i^r = 1 >$ and $P(z_0^r = 1|D^r, \Theta) =< z_0^r = 1 >$. This is equivalent to setting $Q(Z)$ to the current estimate of the posterior distribution.

3. Use these values of $< z_i^r = 1 >$ and $< z_0^r = 1 >$ to set h_t and γ according to equations 4.17 and 4.18.

4. Repeat steps 2 and 3 until the values of h_t and γ are suitably converged.

In the process of finding the optimal values for the parameters we have to calculate the 'responsibilities'. These are $P(z_i^r = 1|D^r, \Theta)$, i.e. how responsible a given object is understood to be for a given source. Once the system has reached equilibrium, after a few iterations of the EM algorithm, these responsibilities give the final estimate of which optical object is the source (i.e. the object with the maximum responsibility) or, if $P(z_0^r = 1|D^r, \Theta)$ is the maximum responsibility, then we infer that the source has no optical counterpart.

4.3 Associations for the N1 FIRBACK Sources Using the Likelihood Ratio and EM Algorithm

The EM algorithm was implemented to find IDs for the 103 175 μm ISO FIRBACK sources in the ELAIS N1 field in the INT–WFS optical catalogue, in order that the IDs and the effectiveness of the method could be directly compared with the likelihood ratio (see Chapter 2). It was halted when the log likelihood converged to less than 10^{-5} per cent with each iteration. This typically took less than 100 iterations.

The EM algorithm was run on all objects in the source error ellipse (i.e. within 90 arcsec) using a bin width for the background magnitude histogram $[P(m_i^r|z_i^r = 0)]$ and the learned true optical counterpart histogram $[P(m_i^r|z_i^r = 1)]$ of $\omega = 1$ magnitude and $\sigma = 20$ arcsec. It was found that the initial estimate for $P(m_i^r|z_i^r = 1)$ and γ did not affect the resultant learned values, after the log likelihood had converged, and hence the IDs made; of course initialising $\gamma = 1$ will ensure all $< z_i^r = 1 >= 0$ and therefore no IDs will be made. The final learned probability distribution of the true optical counterpart magnitudes is shown in Figure 4.1

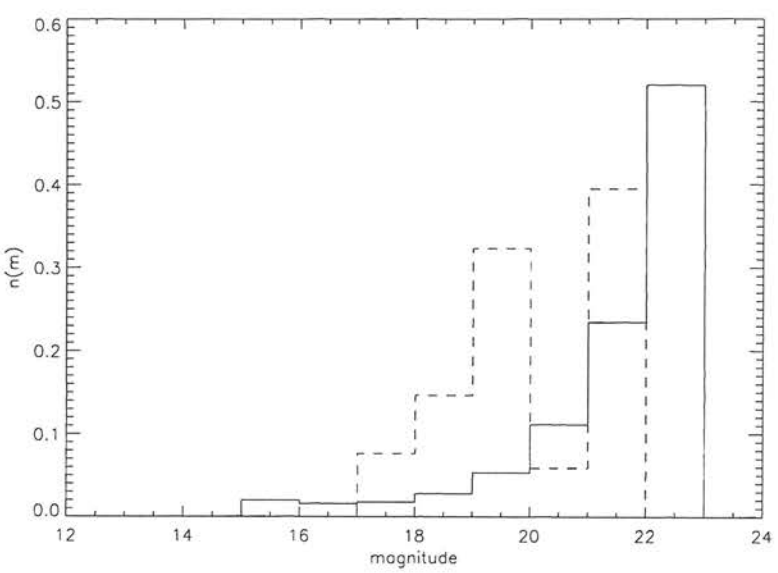

Figure 4.1: Magnitude probability distributions for background objects (solid line) and N1 FIRBACK source IDs (dashed line).

In this case the EM algorithm was able to make IDs for all of the 103 sources, i.e. γ converged to 0.0. For these IDs there was 46 per cent agreement with the IDs found using the likelihood ratio method. Figures 4.2 and 4.3 show a comparison between the

magnitudes and source/ID distances for the IDs made by the EM algorithm and those made by the likelihood ratio where the methods disagree. The EM algorithm failed to identify the brighter likelihood ratio IDs.

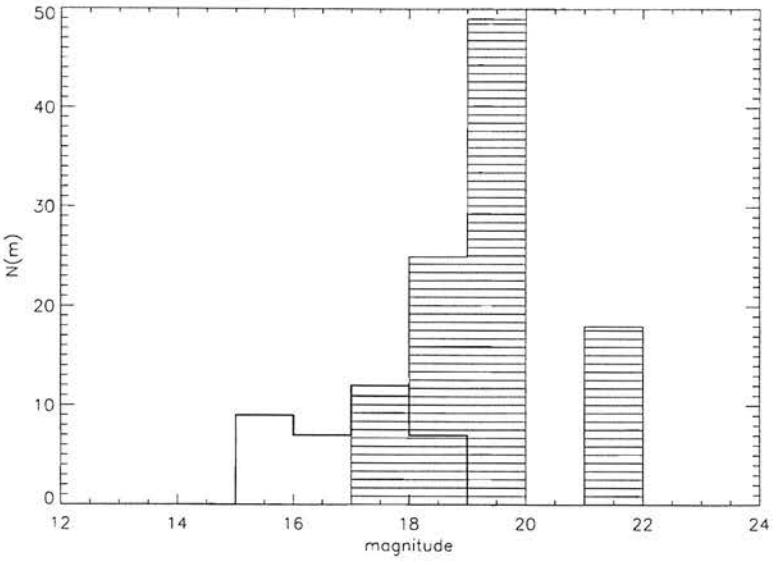

Figure 4.2: Magnitude distribution of the EM N1 FIRBACK IDs (hatched) and magnitude distribution for the likelihood ratio N1 FIRBACK IDs where the EM ID differed from the likelihood ratio ID (thick line).

It was proposed that this could be due to the learned probability distribution for the true counterparts converging to ~ 0.0 in the bright magnitude bins, preventing IDs being selected from these bins, since $< z_i^r = 1 > \propto \sum_{t=1}^{B} h_t T_t(m_i^r)$. In an attempt to avoid this problem a minimum value for the learned histogram bins was set and also a wider bin width of $\omega = 3$ magnitudes tested, however, the EM algorithm still selected fainter IDs than the likelihood ratio method. We also attempted to encourage the algorithm to select bright IDs by initialising the probability distribution for the magnitude of the true source counterparts with 0.0 in the faint (21–24) magnitude bin thus preventing selection of faint counterparts. However, there was still the same failure to identify bright counterparts by the EM algorithm.

The counterpart magnitude selection problem can be understood through examination of the learning mechanism. Look at equation 4.14 for $< z_i^r = 1 >$.

$$< z_i^r = 1 > \propto \left(\sum_t h_t T_t(m_i^r) \right) \prod_{j=1, j \neq i}^{N_r} (\mu_r P(m_j^r | z_j^r = 0)) \tag{4.19}$$

Since we take the product of the background magnitude distribution over $j = 1 - N_r$,

$j \neq i$ and multiply by h_t in the ith bin, $< z_i^r = 1 >$ will be large for objects in the magnitude bin where the ratio $h_t/P(m_i^r|z_j^r = 0)$ is large. The new probability distribution of the counterpart magnitudes is then given by equation 4.17

$$h_t \propto \sum_r \sum_{i=1}^{N_r} < z_i^r > T_t(m_i) \tag{4.20}$$

i.e. this will increase in magnitude bins where $< z_i^r = 1 >$ is large. Since h_t is learned iteratively it will increase further in magnitude bins where $h_t/P(m_i^r|z_j^r = 0)$ is large and decrease rapidly in those bins where $h_t/P(m_i^r|z_j^r = 0) < 1$. This will lead to the selection of IDs with magnitudes in bins where $h_t/P(m_i^r|z_j^r = 0)$ is large which, in the case of the FIRBACK sources, was in the fainter magnitude bins. This was also further exacerbated by the small number of sources leading to a rather noisy h_t distribution. The root of the problem was due to the shape of the background magnitude histogram: this is investigated and compared to that for the HIPASS counterpart catalogue in Section 4.7.

Figure 4.4 shows a comparison of the likelihood ratio P_{ran} values of the IDs for the FIRBACK sources where the likelihood ratio and EM algorithm methods agree and disagree. Both the agreeing and disagreeing IDs have a spread in P_{ran} i.e. those which disagree are not less confident likelihood ratio IDs. The EM algorithm calculates a

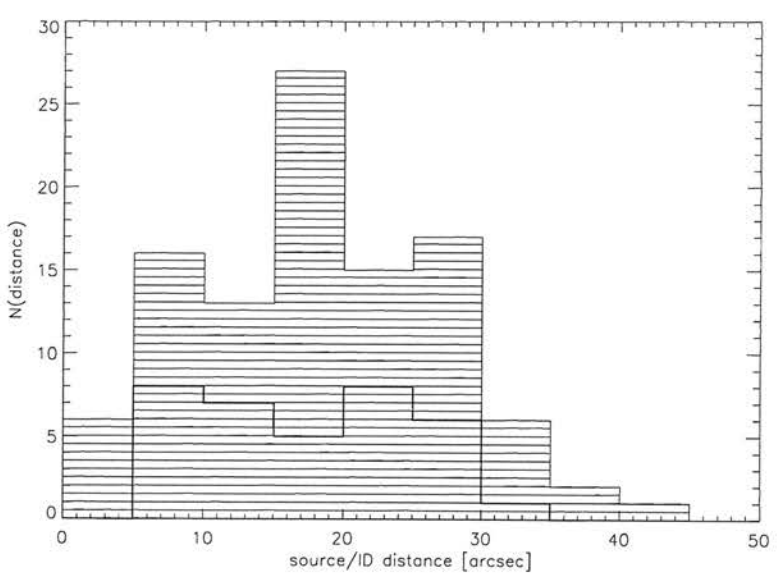

Figure 4.3: Source/ID distance distribution of the EM N1 FIRBACK IDs (hatched) and distribution for the likelihood ratio N1 FIRBACK IDs where the EM ID differed from the likelihood ratio ID (thick line).

responsibility, z_i^r, for each object within the search radius of a particular source and also z_0^r, the probability that the source does not have an optical counterpart. The ID selected for the source is that with the greatest z_i^r; it must also have $z_i^r > z_0^r$. The distribution of z_{ID}^r (i.e. z_i^r values for the chosen IDs) is shown in Figure 4.5. A comparison of the z_{ID}^r values to z_0^r, the probability of the source having no ID, is shown in Figure 4.6. We see from these plots that the IDs selected by the EM algorithm for the N1 FIRBACK sources do not have large z_{ID}^r values; in fact none have $z_{ID}^r > 0.8$, indicating that there may be several objects with $z_i^r > z_0^r$. Figure 4.6 shows that $z_{ID}^r/z_0^r > 10$ for all the selected IDs, therefore, the IDs are many times more likely than the probability of the sources not having IDs. However, all 103 N1 FIRBACK sources were found to have several objects with $z_i^r > z_0^r$ within their error circle: on average there were 40 such objects per source - i.e. all of these objects were more likely to be source IDs than the source not have an ID in the optical catalogue. The distribution of z_{ID}^r/z_2^r (where z_2^r is the responsibility of the second best ID match) was examined; see Figure 4.7. This shows that the probability of these two objects being the ID is very similar: only 30 per cent of source IDs had $z_{ID}^r/z_2^r > 2$. It appears that the algorithm was having difficulty selecting an ID out of many possible objects with $z_i^r > z_0^r$.

z_{ID}^r for the EM IDs was also compared to the likelihood ratio P_{ran} values for instances where they select the same ID (Figure 4.8) and where they select different IDs (Figure 4.9). In both cases there is a spread in P_{ran} and z_{ID}^r. Of the 35 IDs for which the methods disagree, the likelihood ratio ID is the same as another object with $z_i^r > z_0^r$ in 19 cases. Figure 4.10 gives a comparison of the P_{ran} values for the likelihood ratio IDs and those of the second best ID for these 35 sources. There is a spread in the second best P_{ran} values, however, in many cases the P_{ran} of the ID and 2nd best ID are very similar: in 18 cases the second best ID also has a P_{ran} below the cut–off of 0.4 and is therefore also a possible ID in the likelihood ratio method. Figure 4.11 also shows a comparison of the z_{ID}^r and z_i^r for the likelihood ratio chosen ID which has $z_i^r > z_0^r$: we denote this z_{agree}^r. We see that the two objects have very similar responsibility values, out of the 19 sources in the category only 3 have $z_{ID}^r/z_{agree}^r > 2$.

Selecting optical IDs in the INT–WFS catalogue for the N1 FIRBACK sources was a very challenging association problem. Both the likelihood ratio and EM algorithm found it difficult to assign IDs to these sources since there were several possible IDs found by both methods.

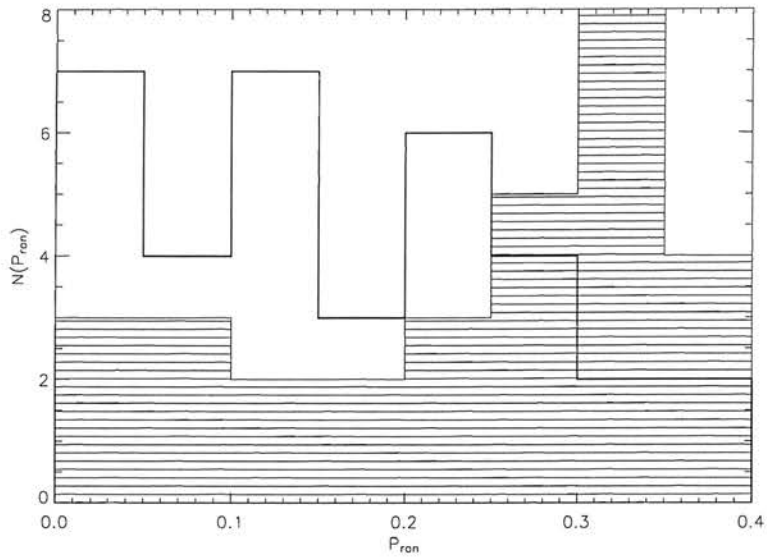

Figure 4.4: Distribution of P_{ran} values for FIRBACK source IDs where the EM algorithm and likelihood ratio methods select the same ID (hatched) and where they select different IDs (thick line).

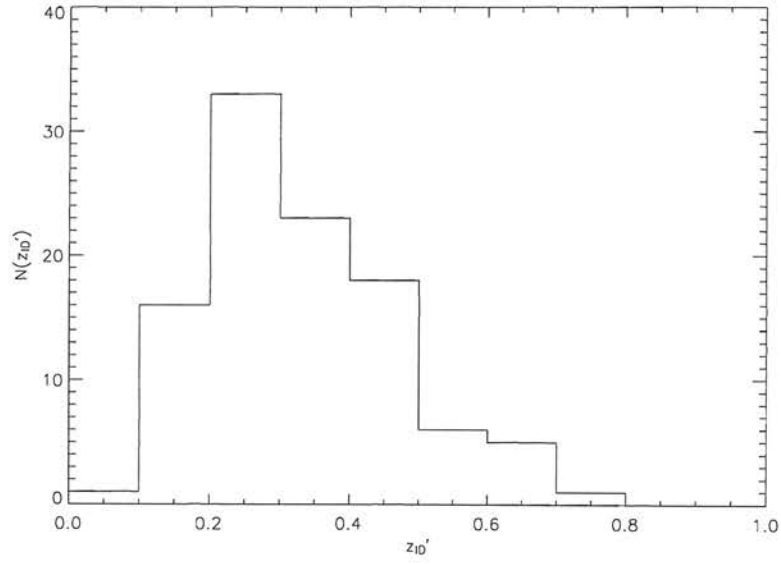

Figure 4.5: Distribution of z^r_{ID} for N1 FIRBACK source IDs selected by the EM algorithm ($\omega = 1$ magnitude).

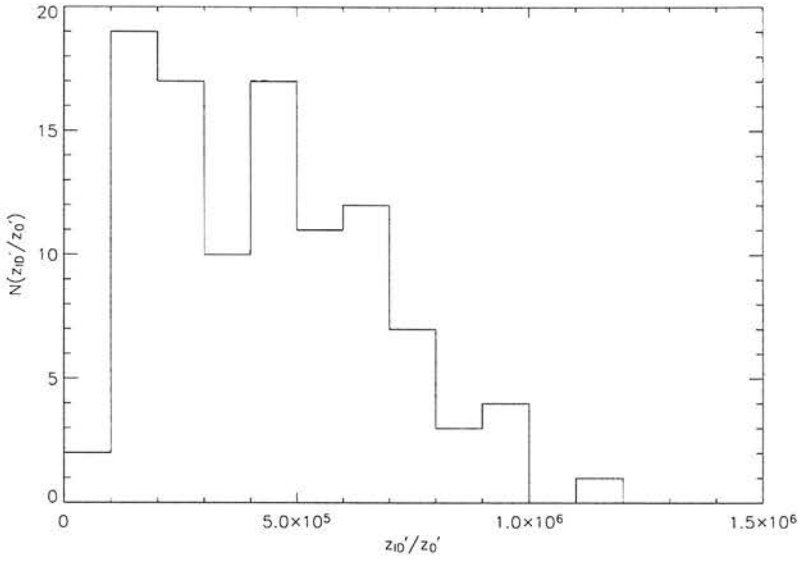

Figure 4.6: Distribution of z^r_{ID}/z^r_0 for the N1 FIRBACK source IDs selected by the EM algorithm ($\omega = 1$ magnitude).

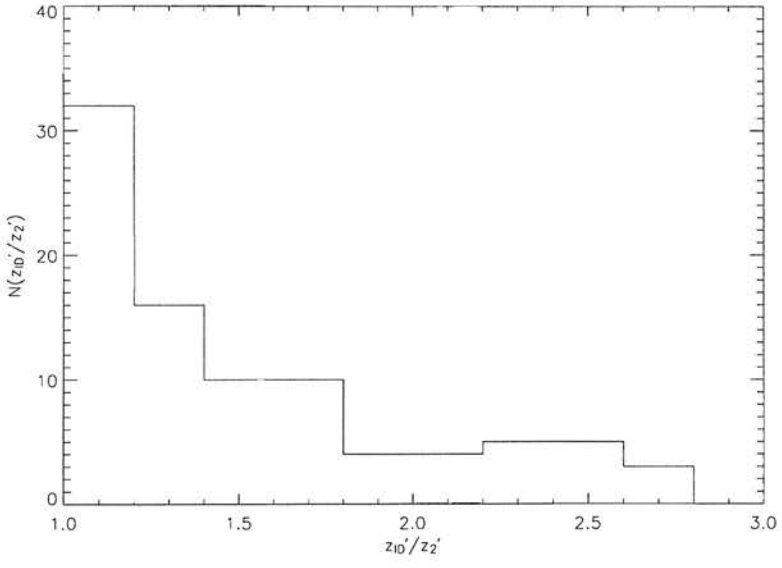

Figure 4.7: Distribution of z^r_{ID}/z^r_2 for all FIRBACK sources assigned an ID by the EM algorithm ($\omega = 1$ magnitude).

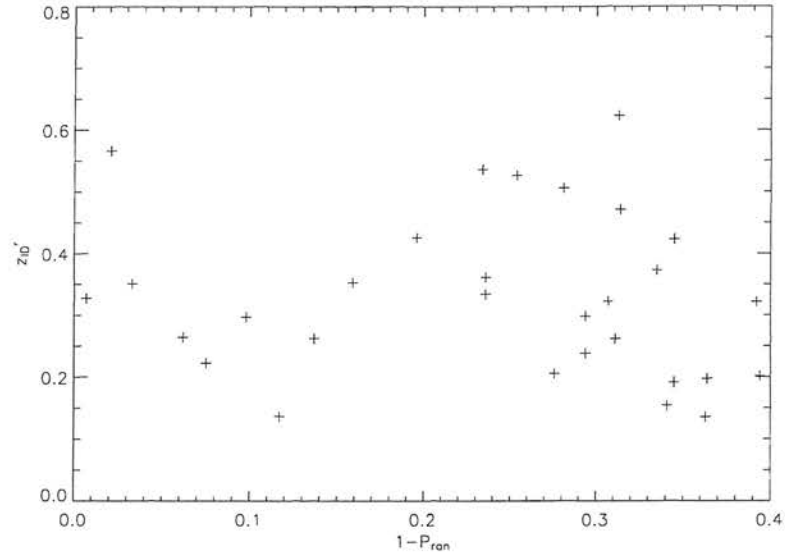

Figure 4.8: z_{ID}^r vs. (1-P_{ran}) for N1 FIRBACK sources where the EM algorithm and likelihood ratio choose the same IDs ($\omega = 1$ magnitude).

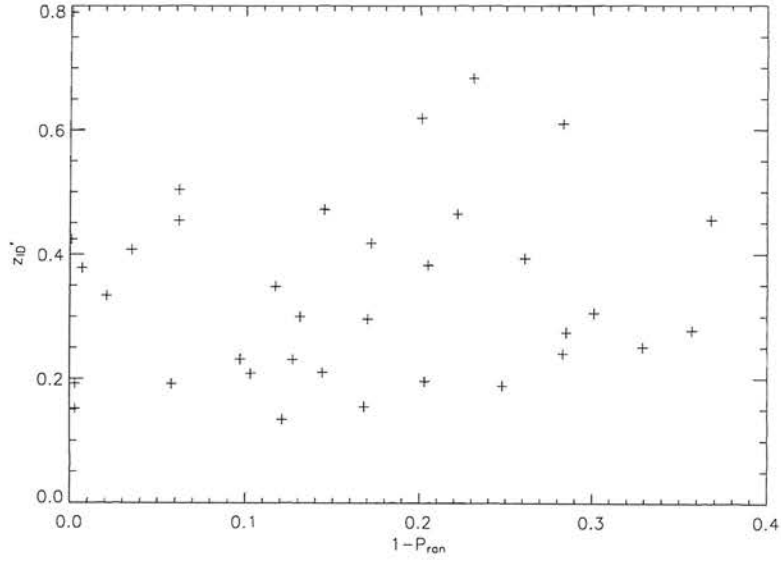

Figure 4.9: z_{ID}^r vs. (1-P_{ran}) for N1 FIRBACK sources where the EM algorithm and likelihood ratio choose different IDs ($\omega = 1$ magnitude).

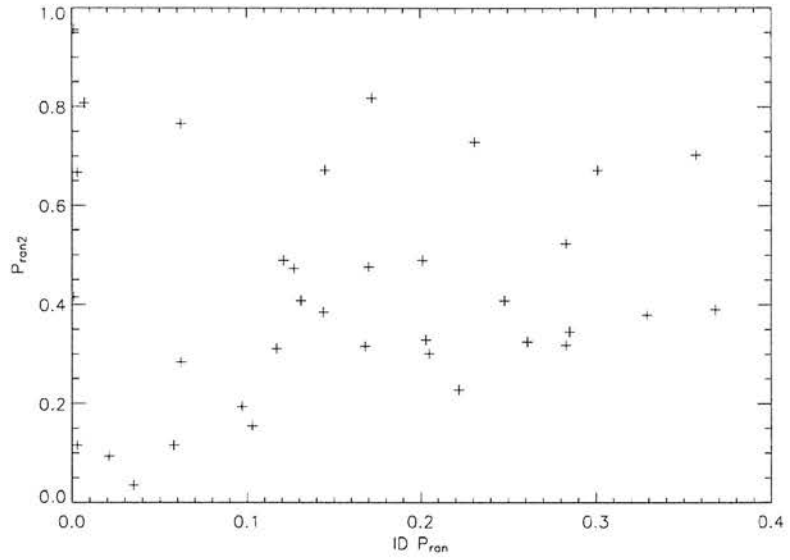

Figure 4.10: A comparison of the P_{ran} values of the ID and second best ID for N1 FIRBACK sources where EM algorithm and likelihood ratio choose different IDs ($\omega = 1$ magnitude).

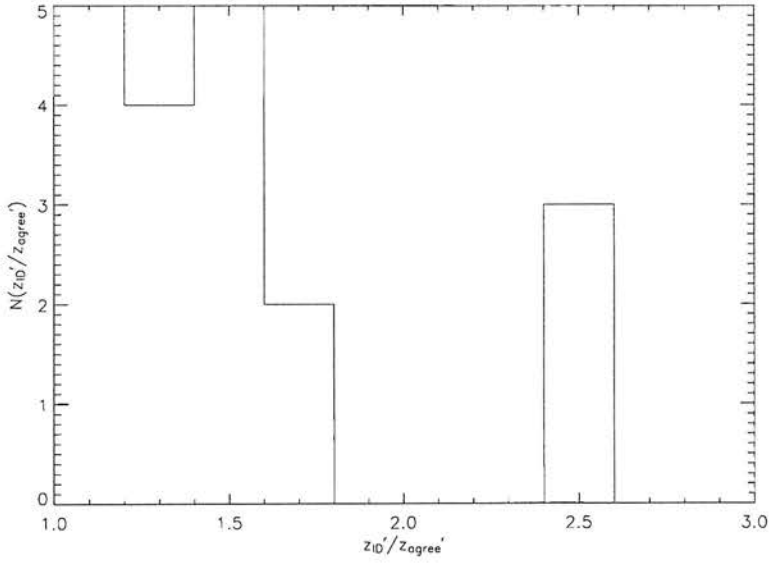

Figure 4.11: Distribution of z^r_{ID}/z^r_{agree} for FIRBACK sources where the likelihood ratio ID agrees with another object with $z^r_i > z^r_0$ ($\omega = 1$ magnitude).

4.4 Using Machine Learning Methods to Obtain HIPASS Associations

The EM algorithm as an association tool was also tested on sources from the HIPASS survey to find the optical IDs in the SuperCOSMOS optical archives: these data sets are described in Sections 4.4.1 and 4.4.2 respectively. Using these data had two advantages. Firstly many of the correct optical IDs were already known via spectroscopic identification and listed in HOPCAT, a catalogue of HIPASS sources with optical IDs found by Doyle et al. (2005) (described in Section 4.4.1), therefore there was a direct test of the IDs made by the EM algorithm. Secondly another machine learning technique for source association had been tested on these data. Rohde et al. (2005) had used a Support Vector Machine (SVM) to find optical IDs for the HIPASS radio sources and kindly allowed us to use their data and directly compare our EM IDs to theirs. This enabled us to assess the relative performance of the two machine learning methods. The SVM and Rohde et al. (2005) results are described in Section 4.4.4 and a comparison of the EM results with those from the SVM in Section 4.5.

4.4.1 The HIPASS Survey

The HI Parkes All–Sky Survey (HIPASS) was a blind survey of the entire sky $\delta < +25°$. It was carried out using the Parkes 64m radio telescope using the multibeam receiver and a hexagonal array of 13 circular feed–horns installed at the focal plane of the telescope. The survey took place from February 1997 to March 2000 for the southern portion ($\delta < +2°$) and until December 2001 for the northern extension ($\delta < +25°$). The telescope scanned across the sky in 8° strips at a rate of $1°\text{mm}^{-1}$ with scans separated by 7 arcmin in RA leading to an effective integration time of 450 s beam^{-1}. The multibeam correlator was used to obtain spectra with a 64MHz bandwidth and 1024 channel configuration, giving a velocity range of -1280 to 12700kms^{-1} and a channel separation of 13.2kms^{-1} at z=0. The HIPASS catalogue (HICAT, Meyer et al. 2004) comprises data from the southern region $\delta < +2°$. It was compiled using detections generated by automatic finder scripts which were then manually verified. HICAT contains 4315 sources for which there are data on their positions, velocities, velocity widths, integrated fluxes and peak flux densities. These sources have positional uncertainties of 0.78 arcmin in RA.

IDs for the HICAT radio sources were manually sought by Doyle et al. (2005) to produce the HOPCAT catalogue, which matched 1878 of the 4315 sources with optical

data. This was compiled by joining the optical candidates to redshifts taken from the Six Degree Field survey (6dF, Wakamatsu et al. 2003) and the NASA Extragalactic Database[1] (NED). If all of the possible optical candidates had redshift information and there was exactly one matching the HICAT redshift then this was accepted as an ID match.

4.4.2 The SuperCOSMOS Sky Survey

The SuperCOSMOS Sky Survey (SSS) (Hambly et al. 2001) aims to digitize the sky in three colours (R, B, I), with R at two epochs, via automatic scans of sky atlas survey photographs. The source photographic material comes from the UK (1.2m), Palomar-I Oschin (1.2m) and ESO (1.0m) Schmidt telescopes which have been systematically photographing the whole sky over the last 50 years. The SuperCOSMOS machine, located in Edinburgh, is a fast, high–precision plate–scanning facility with 0.67 arcsec pixels, 15–bit digitization with accurate positional capability and has been systematically digitizing the Schmidt telescope data. Scans have been completed for the whole southern sky and B and R for the northern sky.

SExtractor (Bertin and Arnouts 1996) was used to analyse images of the SSS regions surrounding the HICAT sources positions. This produced a list of possible optical counterparts for each source. We used the RA, Dec and best magnitude given by SExtractor on the blue scanned optical plates. Although SExtractor returns a stellarity index Rohde et al. (2005) also did their own star/galaxy classification, the results of which we used. They first used software written using the cfitsio library to measure diffraction spikes on all objects. A support vector machine (described in Section 4.4.3) was then trained on a set of 1000 stars and 1000 galaxies using all of the SuperCOSMOS features plus the diffraction spike information. Using a cross validation method (described in Section 4.4.4) a performance of 88 per cent was attained.

4.4.3 Support Vector Machines and Neural Networks

Support Vector Machines (SVM) and Neural Networks are both supervised learning methods which use sets of training data with known classifications to learn about how to classify input data. For example, if we have a model $Y = f(X) + \varepsilon$, where Y is the output, X the input and ε the errors, we can assemble a set of training data $T(x_i, y_i)$, $i = 1, ..., N$ for which, given the inputs, x_i, we know the outputs, y_i. These

[1] http://nedwww.ipac.caltech.edu/

observed inputs, x_i, can be fed into a learning algorithm which will produce an estimate of the outputs $\hat{f}(x_i)$. In response to the differences $y_i - \hat{f}(x_i)$ the learning algorithm can improve its input/output relationship, $\hat{f}$. Once this iterative learning process is complete the outputs $\hat{f}(x_i)$ of the algorithm should be sufficiently close to the true values y_i that the algorithm can be used on further sets of data for which the true outputs are unknown.

If we have a set of data that we wish to separate into different classes it may be possible to find a linear decision boundary or 'hyperplane' in the parameter or 'feature' space of the data, in which case $\hat{f}$ would be a linear function. An optimal separating hyperplane separates these classes by maximising the distance to the closest point from either class, see Figure 4.12.

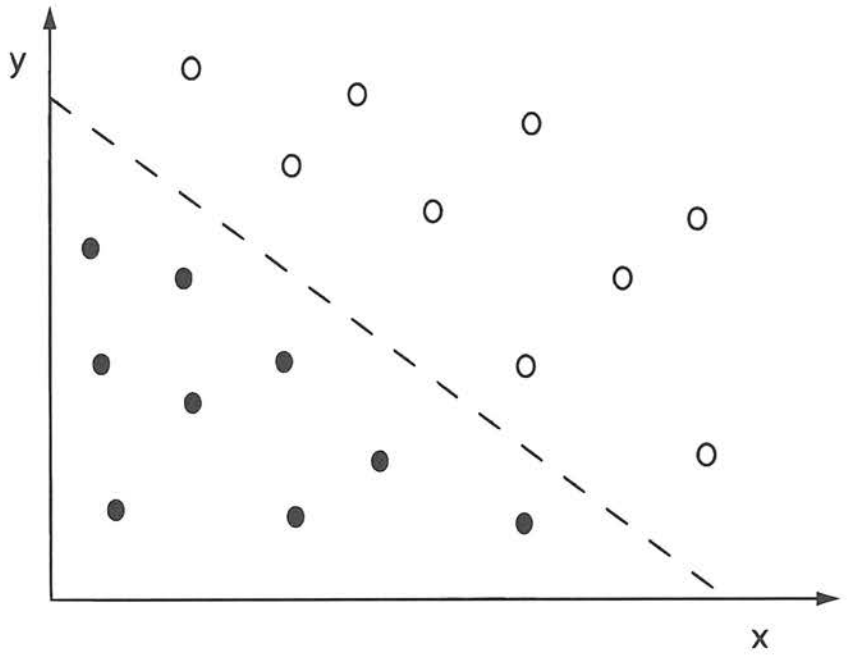

Figure 4.12: An example of two classes (filled circles and open circles) separable by a hyperplane (dashed line).

However, in many cases the classes are not linearly separable; they overlap and a nonlinear boundary must be constructed. The Support Vector Machine achieves this by finding an optimal separating hyperplane (i.e. a linear boundary) in a higher dimensional feature space which is non–linear in the input feature space. It is not necessary to specify the transformation of the input feature vectors to the higher dimensional space directly, simply a kernel function which is the inner product of the transformed feature vectors. The form of the kernel function may be selected by the operator and

many different ones tried to discover that with optimal performance. Popular kernels include linear and polynomial functions and radial basis functions (RBF). RBFs treat the kernel functions as basis functions, these represent a function through an expansion of the transformed inputs i.e.

$$f(x) = \sum_{j=1}^{m} \beta_j h_j(x) \tag{4.21}$$

where $h_j(x)$ are the transformed inputs and β_j are the coefficients, also known as the bias in machine learning. The trade off between smooth and overly complex functions can also be adjusted via a soft margin parameter. This prevents 'overfitting' which describes functions which represent the data well but do not generalise to novel examples.

Neural networks work in a slightly different way to the SVM. A neural network is typically a two–stage model which can be represented by a network diagram, see Figure 4.13. This consists of a network of 'units' represented by circles in the diagram separated into three layers.

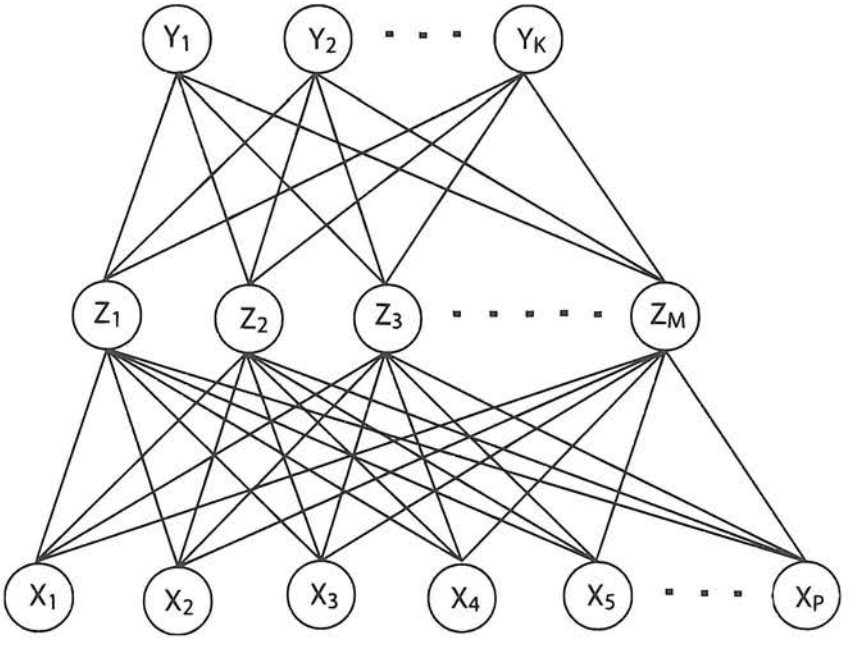

Figure 4.13: Schematic of a single hidden layer, neural network.

For classification of the data into K different classes there are K units at the top of the network with the kth unit modeling the probability of class k. The K target measurements are then Y_k, $k = 1, ...K$: each may take the value 0 or 1 denoting if the datum falls into class k. Z_m are known as the derived features as they are created

from linear combinations of the inputs, X_p, $p = 1, ...P$ and form the central layer of the network. The targets Y_k are then modeled as a function of linear combinations of the derived features Z_m. The units in the middle of the network which compute the derived features are known as hidden units since the values of Z_m are not directly observed and the layer is a 'hidden' layer. A neural network without any hidden units is known as a perceptron.

4.4.4 Support Vector Machines and Neural Networks as a Matching Tool (Rohde et al. 2005)

Support Vector Machines (SVM) and Neural Networks were tested as an astronomical source matching tool by Rohde et al. (2005) on the HIPASS sources with the SuperCOSMOS optical data. These methods were both able to 'learn' about the properties of HIPASS–optical ID pairs from a training set of data which has already been linked (HOPCAT). Their methods and results are briefly described in this section.

First of all the input variables were chosen, i.e. the features. Feature selection is very important to the resulting algorithm performance. In most cases it is best to have a small number of informative features for optimal performance. It may also be advantageous to do some preprocessing of the input variables in situations such as that where the distribution of a variable is not uniform and it may be preferable to transform it prior to 'learning'. Once feature selection had been made a number of algorithms were tested by Rohde et al. (2005) on the training data set and their generalisation performance compared via 10–fold cross validation.

The features used are shown in Table 4.1. These were chosen by Rohde et al. (2005) since it is expected that log(area) and B_j(mag) should be correlated with log(peak flux) and log(integrated flux). There should also be an inverse relationship between velocity and area or magnitude and highly elliptical objects are expected to be linked with radio sources with high velocity widths. Obviously objects that are close to a radio source are expected to be more likely to be matches.

Counterparts for the HIPASS radio sources were sought within 5 arcmin of the source position. Optical objects with areas <600 pixels were discarded in order to avoid the inclusion of massive numbers of small and faint objects. Each radio source/possible optical counterpart pair was assigned a vector which was given a 'positive label' if they were a truly matching pair and a 'negative label' if they did not match. This produced a set of positive training data and a set of negative training data. Both SVMs and neural networks were tested on the training sets and then the one which performed

Table 4.1: Features selected for machine learning inputs

Feature	Origin	Name
1	radio–optical	separation
2	radio	velocity
3	radio	velocity width
4	radio	log[integrated flux (S_{int})]
5	radio	log[peak flux (S_p)]
6	optical	log[isophotal area]
7	optical	semi–major axis
8	optical	semi–minor axis
9	optical	B_j magnitude
10	optical	$B_j - R$
11	optical	$B_j - I$
12	optical	star/galaxy classification

best was used on the unmatched data.

A 10–fold cross–validation test was used to the measure generalisation ability of the implementation of each of the algorithms to discover which algorithm with which parameters performed best. This involves splitting the training data into 10 equal parts, training it on 9 of these and testing it on the 10th. This procedure is repeated 10 times in order to average the result over the entire dataset. This is known as cross validation. The model with the best generalisation was then selected for use on the data for which IDs were as yet unknown. Results of the 10–fold test obtained by Rohde et al. (2005) on the tested algorithms can be seen in Table 4.2. An SVM with a 3rd degree polynomial kernel and soft margin of 10 performed best: overall it was able to identify correctly 99.12 per cent of the positive and negative training data.

Tests were then done to determine the amount of information being provided by the different inputs. This is known as feature importance. It was found that using the optical parameters alone, the SVM was still able to obtain 94 per cent overall correct classification. Rohde et al. (2005) also produced a mismatched data set to test if the radio data were contributing any useful information to the learning process. This consisted of the normal positive matches and a random sample of radio sources matched to distant optical sources (note that the separation parameter was removed from this test). Without the radio data only 47 per cent classification was achieved whereas with the radio information the SVM achieved 72 per cent classification. Therefore it was concluded that there are relationships between the radio and optical properties of these

Table 4.2: Algorithm performances from Rohde et al. (2005). (hu = hidden units, d = order of polynomial)

Algorithm	Soft Margin	Positive Data % correct	Negative Data % correct	Overall % correct
SVM	0.1	87.47 ± 2.24	98.70 ± 0.27	97.17
linear	1	88.94 ± 2.43	98.75 ± 0.50	97.41
	10	88.80 ± 2.44	98.80 ± 0.25	97.43
SVM	0.1	90.04 ± 0.31	99.04 ± 1.67	97.93
polynomial	1	94.18 ± 1.91	99.44 ± 0.20	98.72
d=2	10	96.02 ± 1.47	99.53 ± 0.29	99.05
SVM	0.1	94.91 ± 1.93	99.46 ± 0.27	98.84
polynomial	1	96.24 ± 1.83	99.54 ± 0.20	99.09
d=3	10	96.69 ± 1.26	99.50 ± 0.42	99.12
SVM	0.1	89.39 ± 2.58	99.21 ± 0.27	97.87
RBF=1	1	93.66 ± 2.47	99.50 ± 0.28	98.70
	10	95.43 ± 1.69	99.66 ± 0.17	99.08
Perceptron		86.81 ± 7.73	97.52 ± 2.78	96.05
Neural network	hu=3	93.81 ± 1.68	95.50 ± 1.41	95.27
	hu=4	94.10 ± 2.07	95.46 ± 1.38	95.27
	hu=5	93.50 ± 3.59	95.48 ± 1.26	95.21
	hu=6	93.45 ± 2.01	95.62 ± 1.23	95.32

sources.

Once the SVM with a 3rd degree polynomial kernel and soft margin of 10 had been chosen as the best algorithm it was tested again on the set of sources for which the answer was already known i.e. those with unique velocity matches in HOPCAT. There were 1608 such galaxies and the SVM only misclassified 9. The algorithm could then be tried on the 2221 sources in HOPCAT which did not have IDs. It managed to find 1209 unique matches for these objects, increasing the matched catalogue size to 3096 out of the 4315 sources.

4.4.5 Associations for the HIPASS sources using the Likelihood Ratio Method and the EM Algorithm

As described in Section 4.4.1, the HIPASS sources have a large positional uncertainty. We therefore used a search radius of 5 arcmin within which to search for possible optical counterparts, matching the Rohde et al. (2005) search radius. The positional accuracy of the HIPASS sources was determined from synthetic sources by Zwaan et al. (2004)

and found to be 1.05 arcmin in RA and 0.82 arcmin in Dec, this difference is due to the data being more regularly sampled in the Dec direction. The positional accuracy also agrees very well with results from HICAT matching with the 2MASS Extended Source Catalogue (Jarret et al. 2003). Due to this large positional error a Gaussian width of $\sigma = 100$ arcsec for the spatial distribution of the optical counterparts was used, the optical catalogue also had a completeness limit of -13.5 magnitudes (N.B. these are unnormalised instrumental magnitudes: add 33.4 for a good approximation of the true magnitude).

The EM algorithm was tested on the HIPASS data using a bin width of $\omega = 1$ magnitude for the magnitude probability distributions $P(m_i^r|z_i^r = 0)$ and $P(m_i^r|z_i^r = 1)$. The learned probability distribution for the magnitudes of the true counterparts is shown in Figure 4.14. The algorithm was able to confidently select counterparts for 86 per cent of the 4312 sources (3 had been discarded due to lack of optical information),while of the 1811 HOPCAT sources with confident velocity confirmed optical counterparts, the EM algorithm identified 96 per cent. We note that 56 sources were assigned velocity confirmed IDs further than 5 arcmin from their source position so these IDs could not be found by the EM algorithm (HOPCAT was compiled using a search radius of 7 arcmin). Figure 4.15 shows a comparison of the magnitudes of the EM algorithm optical IDs and those velocity–confirmed links from HOPCAT which it was not able to identify and Figure 4.16 a comparison of their source/optical ID distances. The unidentified velocity–confirmed links are at a range of distances from the sources and to the slightly fainter end of the magnitude distribution.

However, although the EM algorithm returned almost all the HOPCAT velocity–confirmed IDs, since it was given no star/galaxy class information the algorithm also selected many stellar counterparts for the radio sources (1109, of which 176 were velocity–confirmed HOPCAT IDs). It was suggested that some of these stellar objects may be quasars, which are also strong radio emitters, that had been mis–classified. However, the stellar IDs were examined in NED and only 6 classified as quasars. Figure 4.17 shows the magnitude distribution of the stellar IDs and those which are quasars. We note that the quasars are not the brightest 'stellar' objects but are in the centre of the magnitude range. Therefore, several of these stellar IDs could also be quasars which are as yet unknown, however, the HIPASS survey did not expect to find many since most HIPASS sources are at $cz < 6000$ kms-1. At those distances all AGN and QSOs would be expected to be resolved as extended objects and not seen as classical point source QSOs (M. Drinkwater, private communication). This therefore suggests that

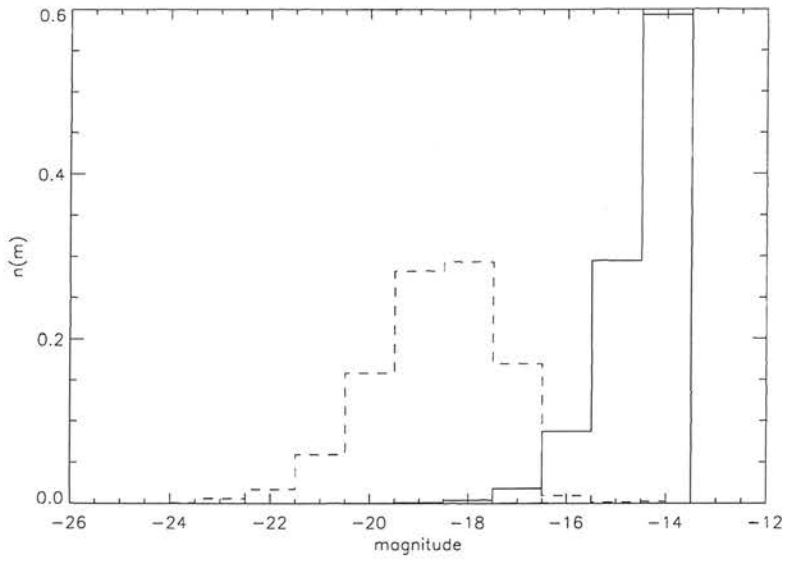

Figure 4.14: Learned probability distribution for instrumental magnitudes of the optical counterparts for the HIPASS sources (dashed line). The magnitude distribution of the background optical catalogue is also given for comparison (solid line).

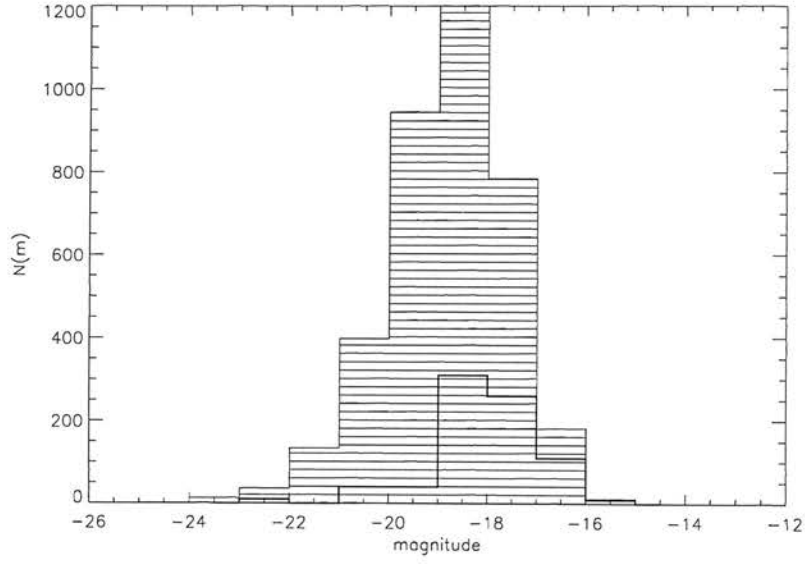

Figure 4.15: Instrumental magnitude histogram for EM HIPASS IDs (hatched) and velocity confirmed HOPCAT IDs not found by EM (thick line, scaled up by a factor of 10).

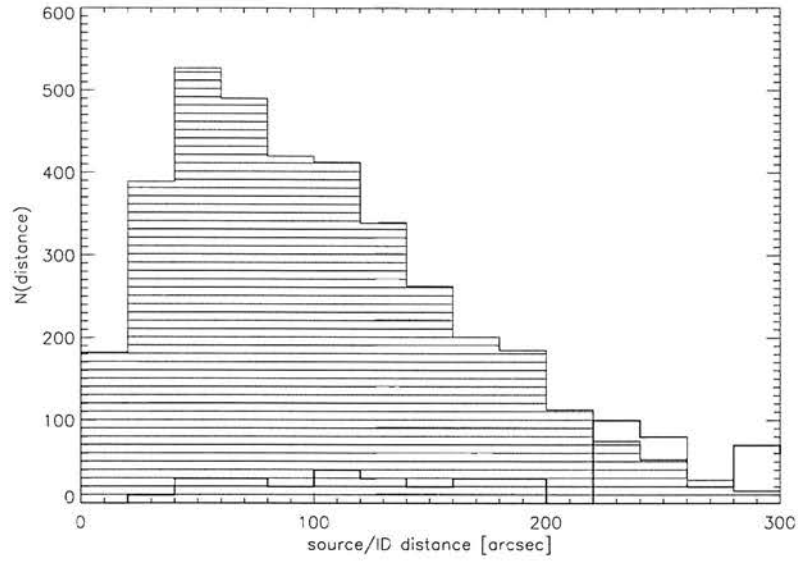

Figure 4.16: Source/counterpart distance histogram for EM HIPASS IDs (hatched) and velocity confirmed HOPCAT IDs not found by EM (thick line, scaled up by a factor of 10).

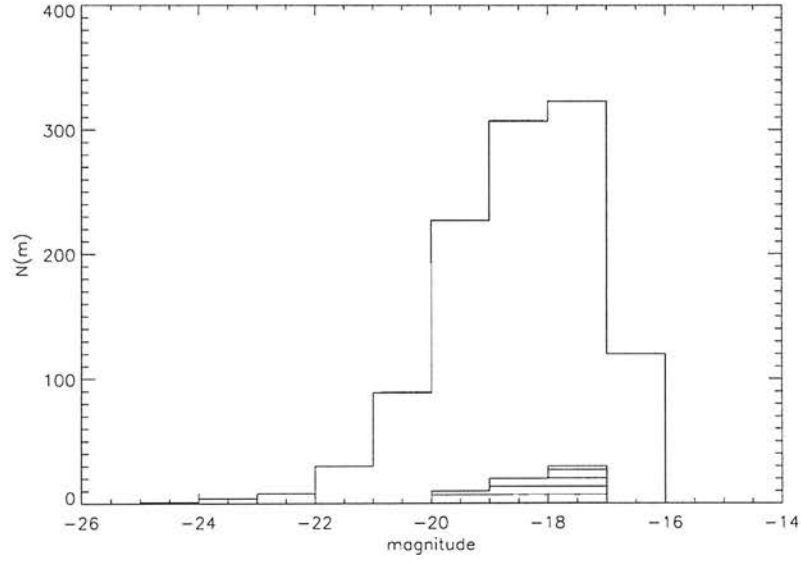

Figure 4.17: A comparison of the instrumental magnitudes of HIPASS source stellar IDs (solid line) and quasar IDs (hatched, scaled up by a factor of 10) selected by the EM algorithm.

some of the EM algorithm selected counterparts may be mis–identifications since we would not expect many of the radio sources to be stellar. The algorithm was therefore run on galaxies only, again using a magnitude bin width of $\omega = 1$ magnitude. 735 of the sources did not have any galaxies within their 5 arcmin error ellipse and therefore could not be assigned optical counterparts, the EM algorithm was able to find IDs for all of the remaining 3577 sources. In this case the algorithm also identified 92 per cent of the HOPCAT velocity–confirmed IDs. However, of the average 172 counterpart candidates within the 5 arcmin search radius per source there was only an average of 1.9 galaxies. This was, therefore, not a challenging learning problem for the galaxy–only EM run, but illustrates why such a large number of stellar IDs were found when no distinction was made between stars and galaxies. This presents an opportunity for future work to incorporate star/galaxy classification into the EM algorithm. It would be interesting to see if the algorithm, when given the star/galaxy classes of the optical objects, could learn the probable star/galaxy classification for the IDs. The large proportion of stars in the southern sky portion of the SSS is due to the inclusion of the Galactic plane. The survey also has a completeness limit of 21.5 magnitudes in the r–band compared to 23 magnitudes of the INT–WFS, used to find counterparts for the FIRBACK survey, it is therefore comparatively shallow and will not contain fainter galaxies at higher redshifts.

The effect of source sample size on the performance of the algorithm was investigated through randomly splitting up the HIPASS catalogue into subsamples of various sizes; from 80 per cent to 3 per cent of its original size (approximately 3450 sources and 135 sources respectively). The results of these EM trials are summarised in Table 4.3. The ability of the algorithm to assign IDs to the sources did not diminish with a reduction in sample size, nor did its ability to identify 'good' IDs i.e. the velocity confirmed HOPCAT IDs. However, there is a greater scatter in the results.

Errors on the learned magnitude probability distribution were estimated using the multiple EM trials on the reduced sized catalogues. Figure 4.18 shows the mean ID magnitude histogram and its associated errors for the 80 per cent–sized catalogues. Errors are very small indicating that the learned magnitude distribution for the source IDs is stable and accurate. The errors on the mean magnitude histogram for source IDs estimated from the 3 per cent–sized catalogues (shown in Figure 4.19) are slightly larger as would be expected. These small catalogues are a similar size to the FIRBACK N1 source list and we infer that the errors on that learned magnitude distribution would be of similar magnitude.

Table 4.3: Performance of the EM algorithm on subsets of the HIPASS catalogue.

No. of sources	% of original catalogue size	% IDs made	% velocity confirmed HOPCAT IDs identified
3450	80	86±1	96±1
2156	50	86±2	96±1
1078	25	85±4	96±3
540	12.5	86±14	95±5
270	6.25	88±12	97±4
135	3	94±20	99±5

The effect of magnitude bin size (ω) on the performance of the EM algorithm was also investigated and found that it did not alter the results greatly. However, if the bin size is reduced dramatically ($\omega < 0.2$) the algorithm overfits the data producing an ID magnitude probability distribution that begins to fit idiosyncrasies in the data.

The Mann et al. (2002) likelihood ratio method of source association was also applied to this data set for comparison. Using the same radius of 5 arcmin within which to search for IDs and a Gaussian width of $\sigma = 100$ arcsec, the likelihood ratio method found IDs for 58 per cent of the HIPASS sources with a $P_{ran} < 0.2$, 69 per cent of sources with $P_{ran} < 0.3$ and 75 per cent with $P_{ran} < 0.4$. The log(LR) and P_{ran} distributions are shown in Figures 4.20 and 4.21 respectively. This method also selected a large number of stellar IDs, 921 of the 3251 IDs with $P_{ran} < 0.4$ and 625 of the 2542 IDs with $P_{ran} < 0.2$.

4.5 Comparison of Associations Made for the HIPASS Sources

The EM algorithm HIPASS IDs were compared to those found using the likelihood ratio using a P_{ran} cut off for good likelihood ratio IDs of both 0.2 and 0.4, which are reasonable if we consider the P_{ran} distribution (Figure 4.21). The likelihood ratio made confident associations for 58 per cent of the radio sources using a cut off of $P_{ran} < 0.2$ and 75 per cent with $P_{ran} < 0.4$ whereas the EM algorithm was able to establish source–ID links for 86 per cent of the HIPASS sources using a magnitude bin width of $\omega = 1$ magnitude. The majority of IDs selected by these two methods were in agreement (2421 source IDs with $P_{ran} < 0.2$ and 2998 with $P_{ran} < 0.4$), however, there were also a number for which they disagreed (121 source IDs $P_{ran} < 0.2$ and 236

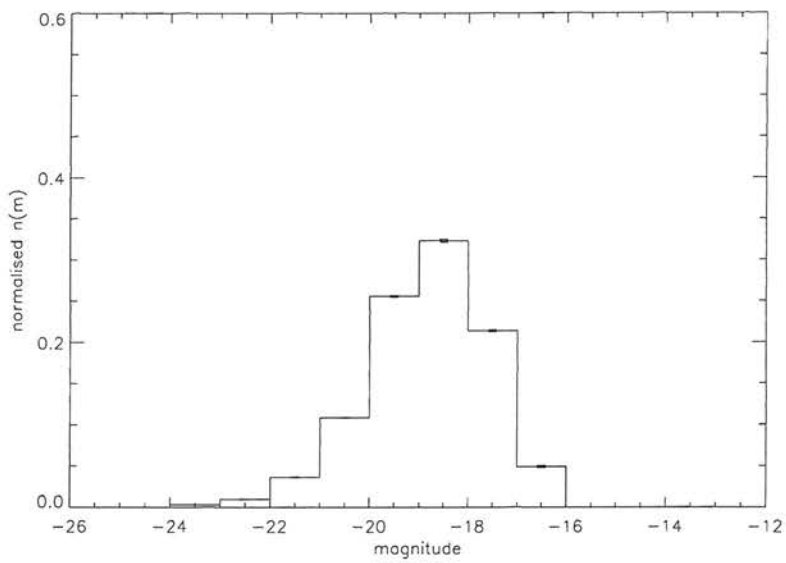

Figure 4.18: Mean instrumental magnitude histogram for EM IDs found for catalogues 80 per cent of the size of the original HIPASS catalogue with associated error bars.

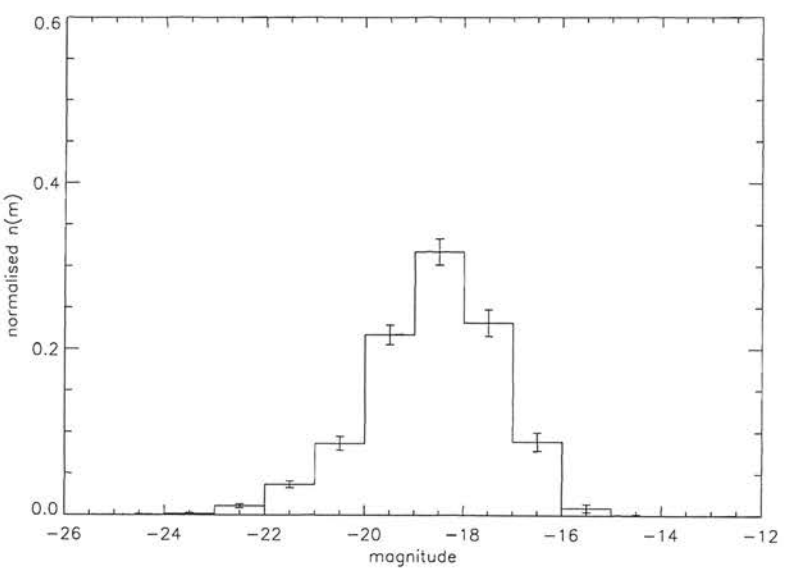

Figure 4.19: Mean instrumental magnitude histogram for EM IDs found for catalogues 3 per cent of the size of the original HIPASS catalogue with associated error bars.

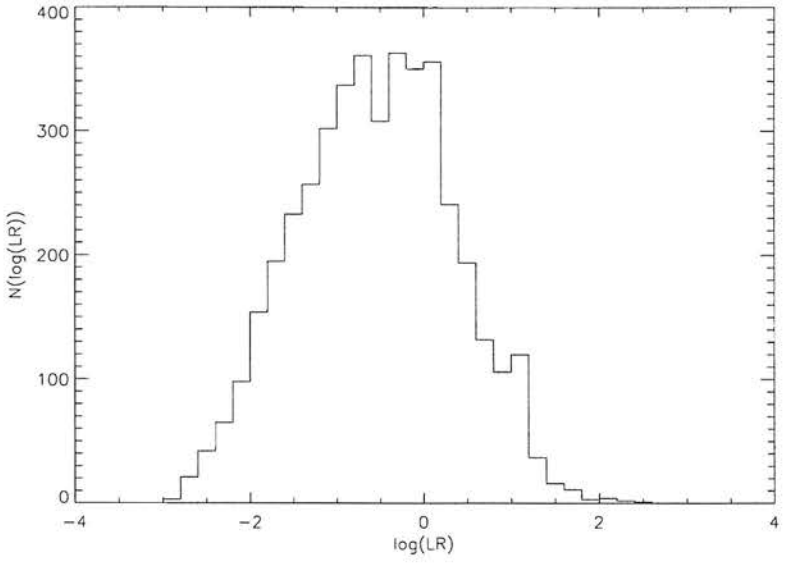

Figure 4.20: log(LR) distribution for likelihood ratio chosen IDs for the HIPASS sources.

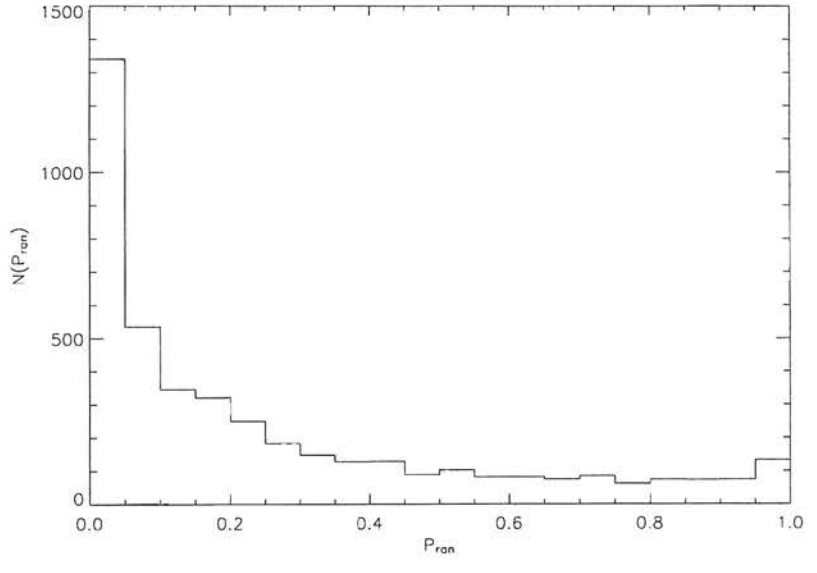

Figure 4.21: P_{ran} distribution for likelihood ratio chosen IDs for the HIPASS sources.

with $P_{ran} < 0.4$). The magnitudes and source–ID distances of these disagreeing IDs are compared in Figures 4.22 and 4.23 for likelihood ratio IDs with $P_{ran} < 0.2$ and Figures 4.24 and 4.25 for IDs with $P_{ran} < 0.4$. The likelihood ratio method appears to select optical IDs which are slightly brighter but further from the source position compared to the EM algorithm IDs. If we look at the learned probability distribution for the instrumental magnitudes of the IDs (Figure 4.14) we see that the probability of the ID having an instrumental magnitude in the range -19.5 to -17.5, corresponding to the peaks of the distributions of instrumental magnitudes of the disagreeing IDs, is very similar. Therefore, for objects in this range the ID will be selected on its distance to the source.

It was also noticed that both the likelihood ratio and EM algorithm methods selected a number of stellar IDs for the radio sources. This was inevitable due to the enormous proportion of stars in the field ($\sim$ 170/172 objects/source). It might be an interesting extension to this work to incorporate into the EM algorithm the learning of the star/galaxy class of the ID as well as its magnitude just as the SVM was able to.

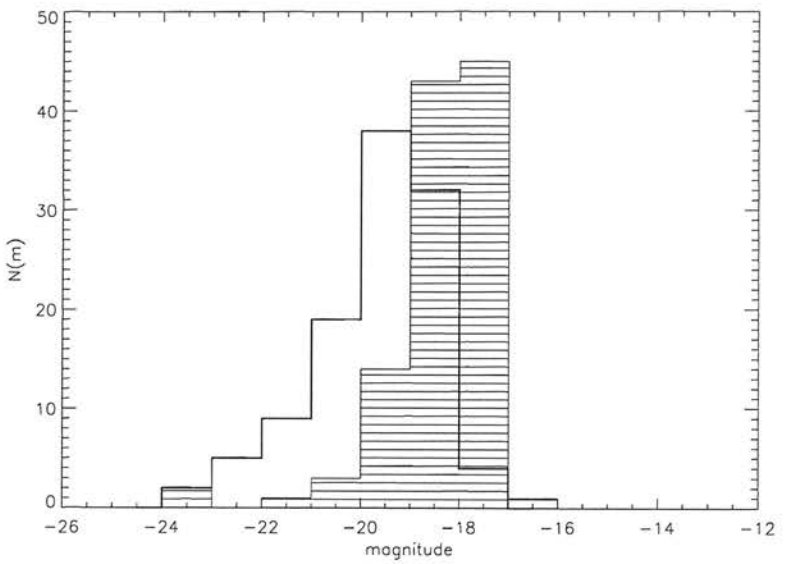

Figure 4.22: Instrumental magnitude histograms for EM HIPASS IDs (hatched) and likelihood ratio IDs ($P_{ran} < 0.2$, bold line) where the two methods disagree on the ID.

The distribution of z^r_{ID} for the EM algorithm selected IDs for the HIPASS sources is given in Figure 4.26. This can be compared to the P_{ran} distribution (Figure 4.21) created by the likelihood ratio method. We note that for the HIPASS sources the

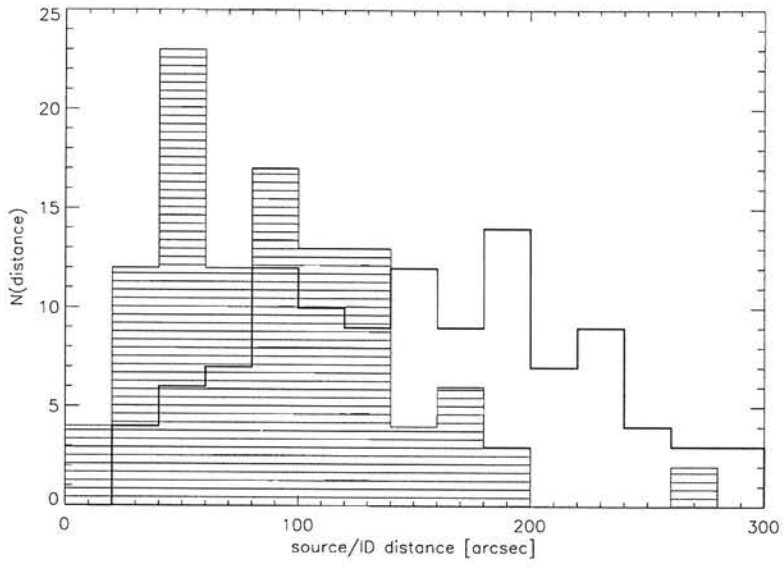

Figure 4.23: Source/counterpart distance histograms for EM HIPASS IDs (hatched) and likelihood ratio IDs ($P_{ran} < 0.2$, bold line) where the two methods disagree on the ID.

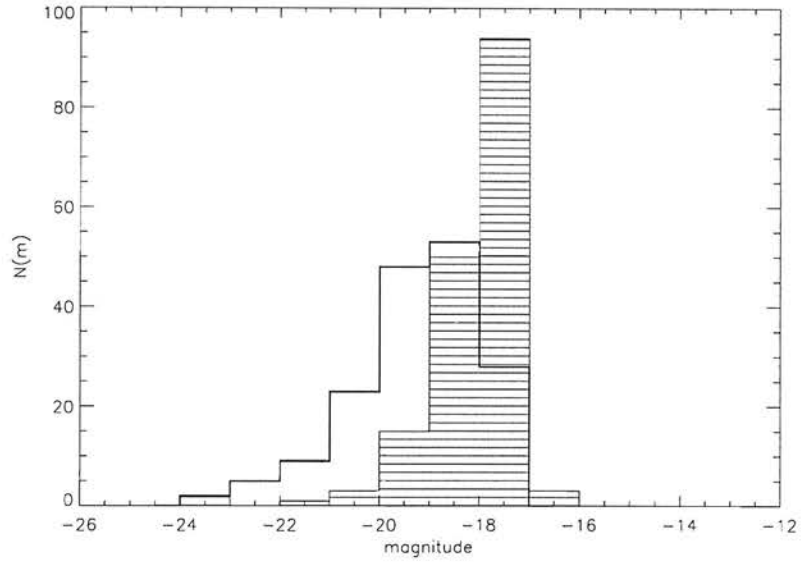

Figure 4.24: Instrumental magnitude histograms for EM HIPASS IDs (hatched) and likelihood ratio IDs ($P_{ran} < 0.4$, bold line) where the two methods disagree on the ID.

z^r_{ID} values are high, unlike for the FIRBACK sources and there are no EM IDs with $z^r_i < 0.1$ since the z^r_0 value explicitly gives the cut–off value for good associations which had to be arbitrarily chosen in the likelihood ratio method. The ratio z^r_{ID}/z^r_0 was also examined (Figure 4.27); 91 per cent of sources had $z^r_{ID}/z^r_0 > 2$ and 78 per cent had $z^r_{ID}/z^r_0 > 5$: i.e. the majority of IDs are at least twice as likely as the source not having an optical ID in the catalogue. However, there were 1893 sources with more than one object within their error ellipses with $z^r_i > z^r_0$. Overall there were 1.8 objects per source with $z^r_i > z^r_0$ i.e. all of these objects were more likely to be the source ID than the source not have an ID in the optical catalogue. The distribution of z^r_{ID}/z^r_2 is shown in Figure 4.28, for some sources the ID is up to 300 times more likely than the next best object and 39 per cent have $z^r_{ID}/z^r_2 > 2$. Where there are two objects with very similar probabilities of being the ID these could be multiple IDs and should both be considered when using the results of this association process.

z^r_{ID} values were compared to the likelihood ratio P_{ran} ID values (using a P_{ran} cut–off of 0.2) for the HIPASS sources for instances where they select the same ID (Figure 4.29) and where they select different IDs (Figure 4.30). We see that when the two methods agree on the source ID there is a clustering of objects with low P_{ran} and large z^r_{ID} indicating that these are reliable IDs in both methods. Where the two methods

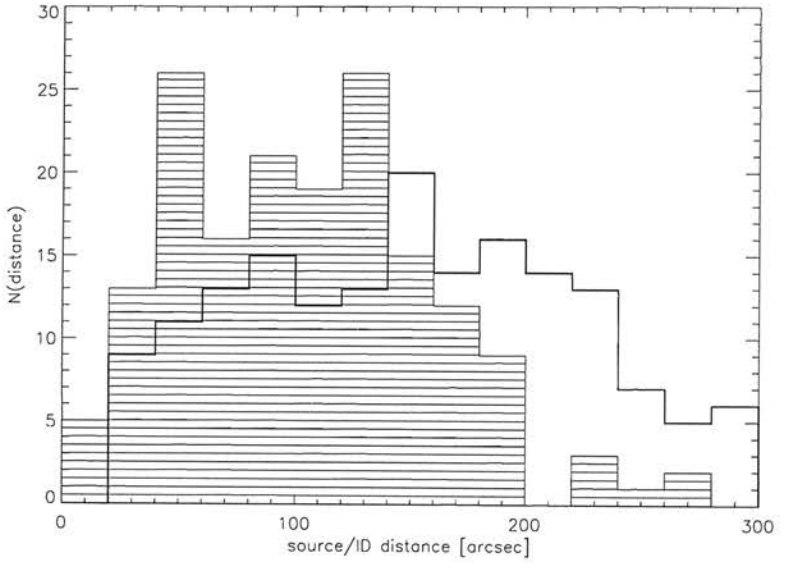

Figure 4.25: Source/counterpart distance histograms for EM HIPASS IDs (hatched) and likelihood ratio IDs ($P_{ran} < 0.4$, bold line) where the two methods disagree on the ID.

select different IDs there is a spread in both P_{ran} and z^r_{ID}. Out of the 109 disagreeing IDs the likelihood ratio ID is the same as another object with $z^r_i > z^r_0$ in 100 cases. The P_{ran} values for these disagreeing IDs were compared to those of the second best likelihood ratio ID for these sources, see Figure 4.31. The majority of the P_{ran} values of the IDs and second best IDs are very similar, in 78 of the 109 cases the second best ID also has a P_{ran} below the cut–off of 0.2. The responsibility values z^r_{ID} and z^r_i of the likelihood ratio IDs which had $z^r_i > z^r_0$ were also compared (Figure 4.32), only 22 out of the 100 cases were found to have $z^r_{ID}/z^r_{agree} > 2$. For this small number of sources both the EM algorithm and likelihood ratio methods had difficulty choosing between two objects for the source ID.

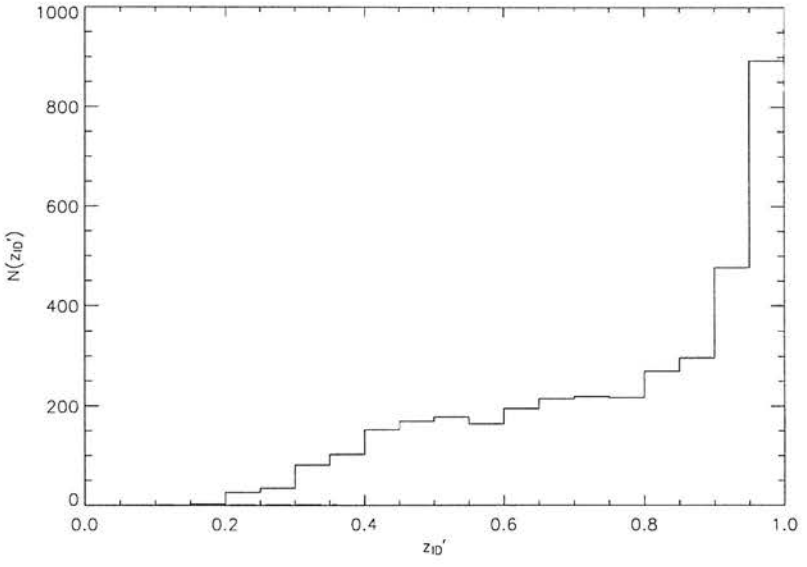

Figure 4.26: Distribution of z^r_{ID} for HIPASS source IDs selected by the EM algorithm.

The support vector machine association method of Rohde et al. (2005) increased the number of IDs for the HIPASS sources from 1878 in HOPCAT to 3096 i.e. 70 per cent of the catalogue. Of these optical associations the EM algorithm method agreed with 95 per cent. The two machine learning association methods, therefore, selected an almost identical optical population for the counterparts of the radio sources.

It is also possible to use the EM algorithm in a supervised manner using a set of training data in analogy with the SVM, by explicitly telling it that a known ID has $z^r_i = 1$. In this case the EM algorithm selected counterparts for 70 per cent of the HIPASS catalogue; a very similar performance to the supervised SVM method.

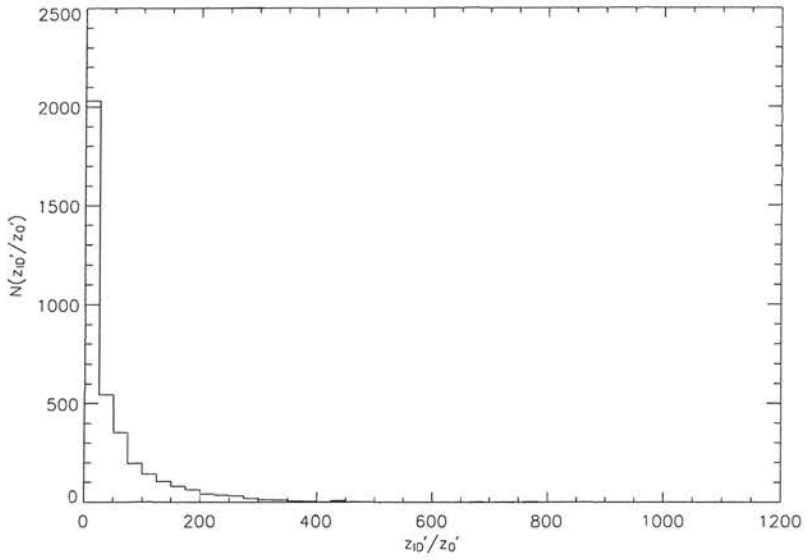

Figure 4.27: Distribution of z^r_{ID}/z^r_0 for the HIPASS sources IDs selected by the EM algorithm.

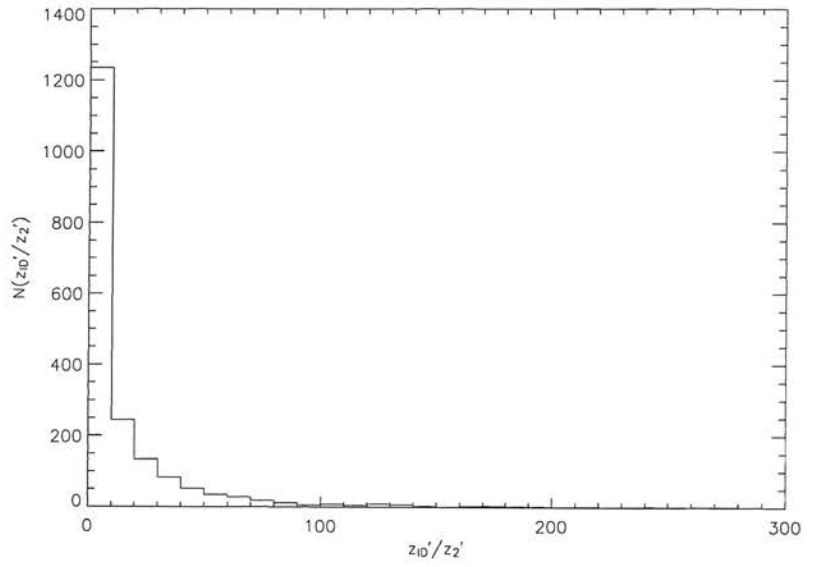

Figure 4.28: Distribution of z^r_{ID}/z^r_2 for all HIPASS sources assigned an ID by the EM algorithm.

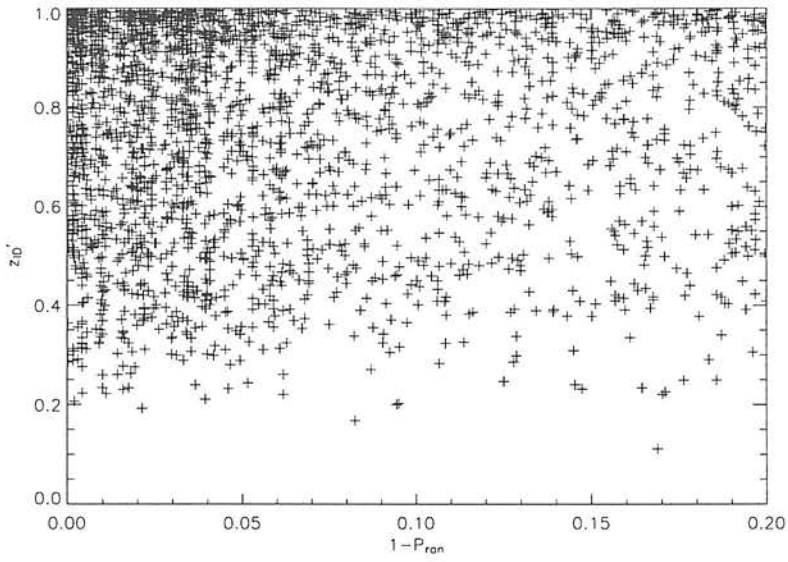

Figure 4.29: z^r_{ID} vs. $(1-P_{ran})$ for sources where the EM algorithm and likelihood ratio choose the same ID for HIPASS sources.

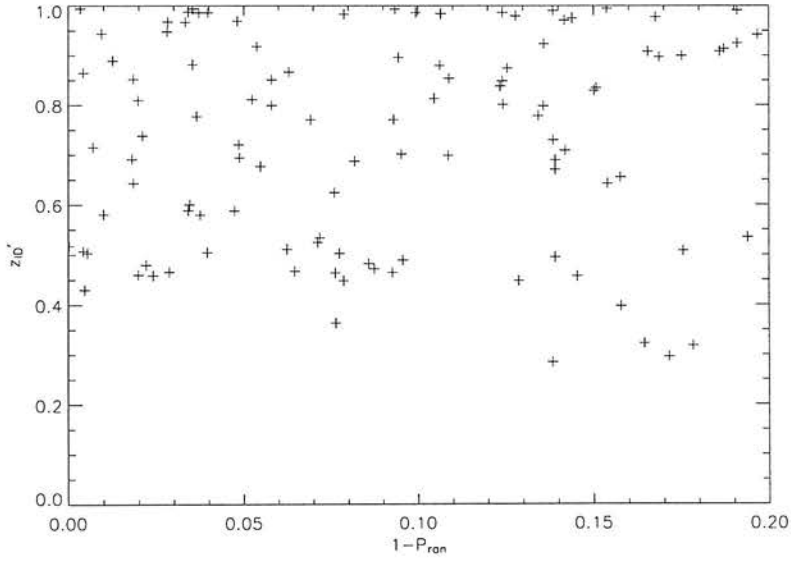

Figure 4.30: z^r_{ID} vs. $(1-P_{ran})$ for sources where the EM algorithm and likelihood ratio choose different IDs for HIPASS sources.

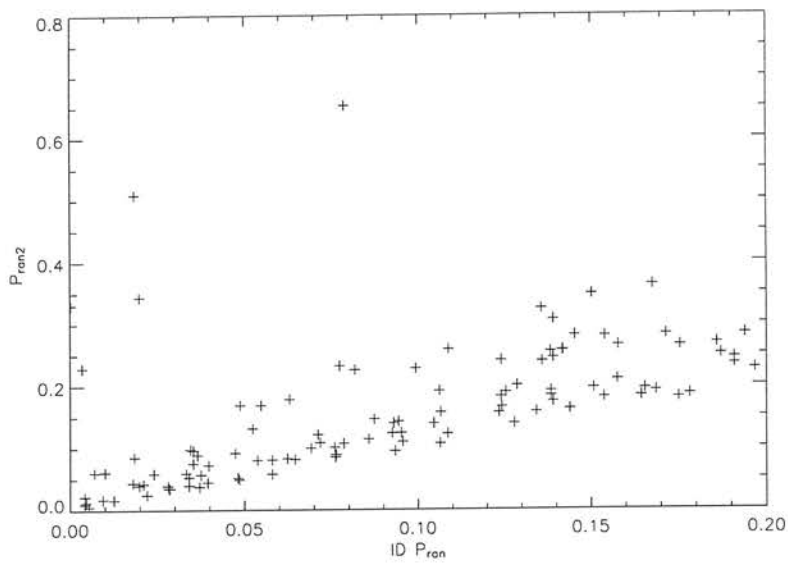

Figure 4.31: A comparison of the P_{ran} values of the ID and second best ID for sources where the EM algorithm and likelihood ratio choose different IDs for HIPASS sources.

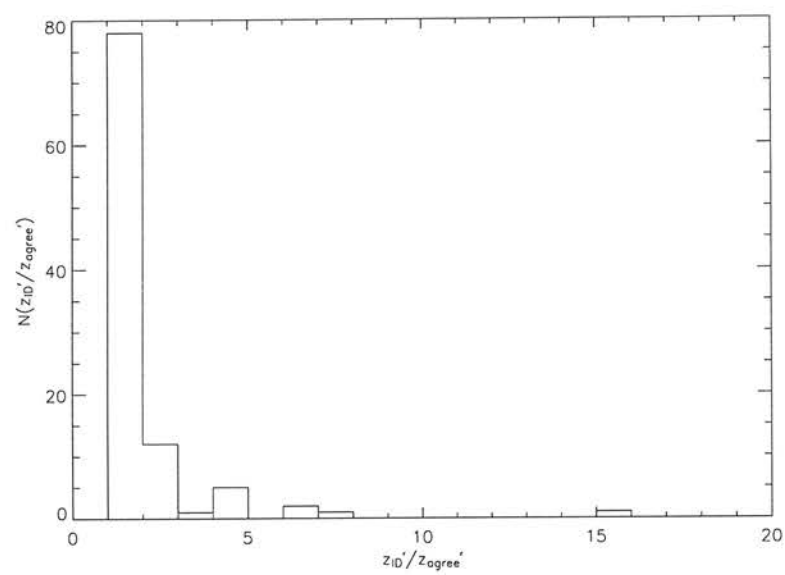

Figure 4.32: Distribution of z_{ID}^r/z_{agree}^r for HIPASS sources where the likelihood ratio ID agrees with another object with $z_i^r > z_0^r$.

4.6 Comparison of the Performance of the EM Algorithm on the Two Datasets

The EM algorithm performed extremely well on the HIPASS dataset identifying counterparts for 86 per cent of the radio sources and correctly selecting 96 per cent of the velocity confirmed HOPCAT IDs. However, it did not perform as well on the much smaller FIRBACK source catalogue. It made IDs for all 103 sources, however, only 46 per cent of these were the same as the likelihood ratio method IDs. The IDs were also fainter than those selected by the likelihood ratio method. It was thought that this might be due to some of the bright magnitude bins in the learned magnitude probability distribution converging to ~ 0.0. To try to prevent this a wider bin width of $\omega = 3$ magnitudes, initialising the distribution with $h_t = 0$ in the faint bin and giving the bins a minimum value were tried but these did not rectify the discrepancy.

Since the HIPASS survey area is so much larger than that of the ELAIS N1 region the background magnitude histogram obviously contains many more objects. It was, therefore, postulated that the background magnitude histogram for the FIRBACK source EM algorithm application might be rather noisy. To test this 10 randomly selected subsamples of the HIPASS survey optical background, of the same size as the FIRBACK N1 optical background, were selected. The resulting magnitude distributions were identical to that of the entire region. Noise in the background magnitude histogram for the FIRBACK sources was therefore not causing the problem.

The problem was in fact due to the learned counterpart magnitude probability distribution growing rapidly in magnitude bins where the ratio $h_t/P(m_i^r|z_i^r = 0)$ was large. In the FIRBACK survey case, this occurred in the fainter magnitude bins and in the bright bins the learned magnitude distribution reduced rapidly (see Section 4.3). Since we did not have the same problem for the HIPASS survey application we compared the background magnitude histograms (Figure 4.33). The background magnitude histogram of the SSS optical data for the HIPASS survey EM algorithm application has a long tail stretching to bright magnitudes, this allows the selection of IDs in these magnitude bins since $h_t/P(m_i^r|z_i^r = 0)$ will be large. However, in the FIRBACK survey case the background INT–WFS magnitude distribution does not have this tail, therefore, $h_t/P(m_i^r|z_i^r = 0)$ will be much less in the bright magnitude bins and IDs will be much less likely to be selected from these bins.

We created model data to simulate background magnitude distributions with different shapes. These varied from those with a long tail at bright magnitudes, such as the

SSA data in the HIPASS case, to those with an overall much flatter distribution, such as the INT–WFS data in the FIRBACK case. Model possible counterpart lists for the 103 N1 FIRBACK sources were constructed, each containing 100 background objects and one true ID. The positions of the background objects were distributed randomly over the 90 arcsec search radius. The true IDs had positions selected from a Gaussian distribution with width of 20 arcsec and magnitudes selected from a uniform distribution over magnitudes 13 to 18. Five model background magnitude distributions were constructed over the magnitude range 13–23, these are shown in Figure 4.38. These range from model A with a long flat tail at bright magnitudes to model E which is a much flatter distribution.

The EM algorithm and likelihood ratio method were tested on the model data, the results are summarised in Table 4.4. Both association methods performed very well on model A and progressively worse through to models D and E where the EM algorithm did not assign any IDs i.e. $z_0^i > z_i^r$ in all cases. The shape of the background magnitude distribution therefore has a great effect on the ability to assign IDs to a list of sources for both methods. In the EM algorithm case this is due to the ratio $h_t/P(m_i^r|z_i^r = 0)$, where this is large the algorithm is able to select IDs but when the background distribution is relatively flat and $h_t/P(m_i^r|z_i^r = 0)$ is small it has difficulty assigning IDs. For the likelihood ratio case $LR \propto e^{-r^2/2\sigma^2}/N(m)$, therefore when the background magnitude distribution has a long bright tail N(m) will take a wide range of values over the magnitude range. However, when the background magnitude distribution histogram is flatter the heights of the magnitude histogram bins will be very similar and therefore the likelihood ratio values of objects also similar making it difficult to select IDs.

The difference in the background magnitude histograms for the two datasets is due to the properties of the two different fields. The four main ELAIS survey regions in which the FIRBACK survey was carried out were selected for their low Galactic cirrus contamination and also positioned at a high ecliptic latitude to avoid zodiacal contamination ($|b| > 40$). These fields are therefore well out of the Galactic plane and will contain few nearby bright stars and, due to their restricted size, few other bright objects, hence, the lack of tail to bright magnitudes in the background magnitude histogram for the N1 region. In contrast the HIPASS survey covered the entire southern sky including the Galactic plane and many bright galaxies. Therefore, the background magnitude histogram of the HIPASS survey region has a long tail to bright magnitudes. We can clearly see the cause of these features if we look at the magnitude histograms for the

stellar objects and galaxies separately. Figure 4.35 shows the stellar and galaxy magnitude histograms for the HIPASS survey region, while Figure 4.36 gives a normalised version. In this field of the average 172 objects within 5 arcmin search radius only 1.9 were galaxies. In the HIPASS region, which covered the entire southern sky, both stars and galaxies have similar shaped magnitude distributions. There are far more faint stars and galaxies than bright ones. This leads to an overall magnitude distribution with a long tail to bright magnitudes and we see that the galaxy population contains relatively more bright objects than the stellar population. Hence the EM algorithm was able to select many galaxies as IDs for the HIPASS sources even though $\sim 170/172$ objects per source were stellar. If we compare the magnitudes of the stellar objects and galaxies chosen as IDs (Figure 4.37) we see that their distributions are very similar. The algorithm learned the magnitude distribution of these true galaxy IDs, however, since it had no way of distinguishing between stellar objects and galaxies the few stellar objects present with such bright magnitudes (see Figure 4.36) were also selected as IDs. Figure 4.34 shows the magnitude histograms for the ELAIS N1 region, in this field there were approximately 72 objects within the 90 arcsec search radius per source of which on average 49 were galaxies. The FIRBACK N1 survey region covered only 2 deg^2 and although the galaxy magnitude histogram is a similar shape to those for the HIPASS region the stellar magnitude distribution is very flat. We suggest that this is due to the fact that the field was selected to contain very few stars, it may also be in part due to the small area being unusual in its number of faint stars. The addition of this flat stellar magnitude histogram to the more typical magnitude distribution produces the overall magnitude distribution with a flat truncated tail to bright magnitudes. As the tests on model data showed, the EM algorithm is more dramatically affected by the background magnitude histogram than the likelihood ratio and had more difficulty assigning IDs to the N1 FIRBACK sources.

4.7 Summary

We implemented a machine learning technique, the EM algorithm, to find optical associations for two different source populations and compared these to those found by the likelihood ratio method and, for one of these populations, another machine learning technique, the Support Vector Machine.

The SVM method requires a set of training data consisting of a number of sources that have already been matched with counterparts i.e. it is a supervised learning

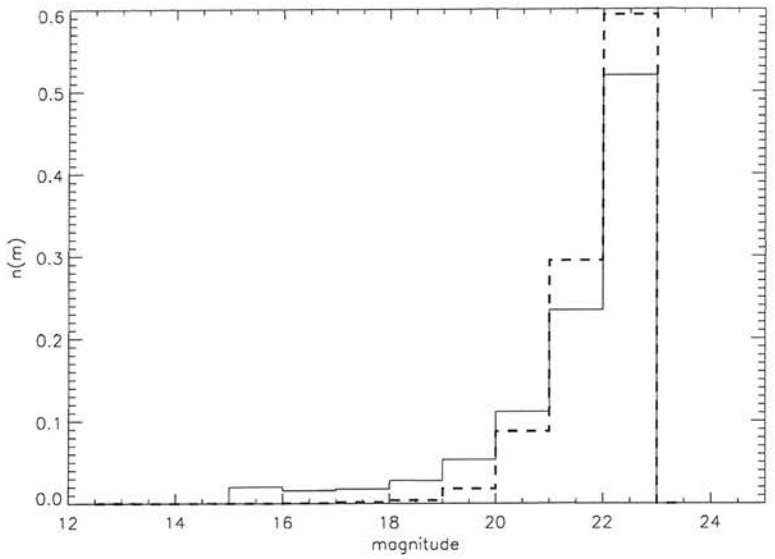

Figure 4.33: A comparison of the background magnitude histograms for the optical catalogues within which IDs were sought for the FIRBACK N1 (solid line) and HIPASS (dashed line) sources. Histograms are normalised to 1 and scaled to the same limiting magnitude to enable comparison.

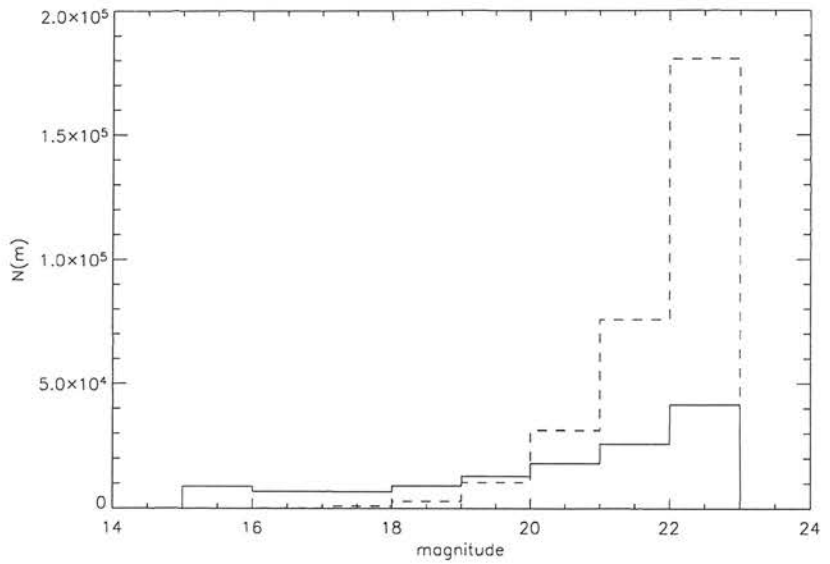

Figure 4.34: A comparison of the background magnitude histograms for stellar objects (solid line) and galaxies (dashed line) within the ELAIS FIRBACK N1 survey region.

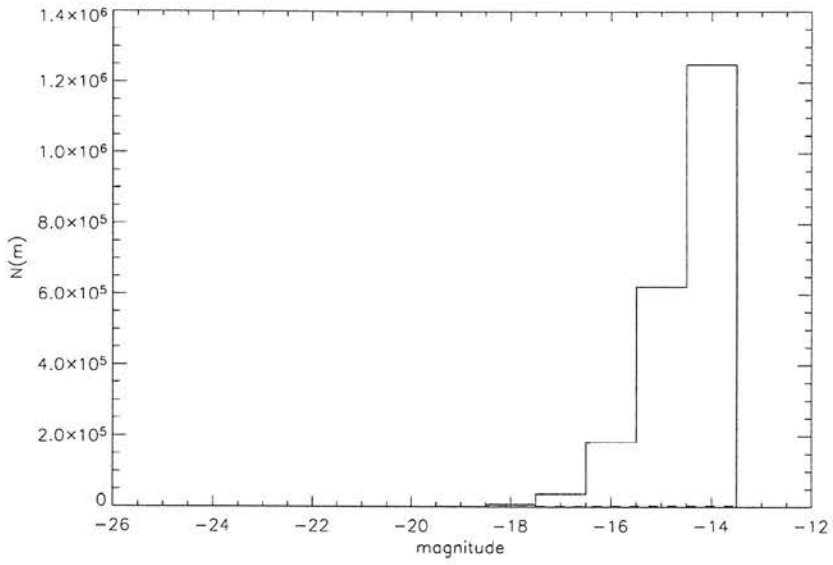

Figure 4.35: A comparison of the background magnitude histograms for stellar objects (solid line) and galaxies (dashed line) within the HIPASS survey region.

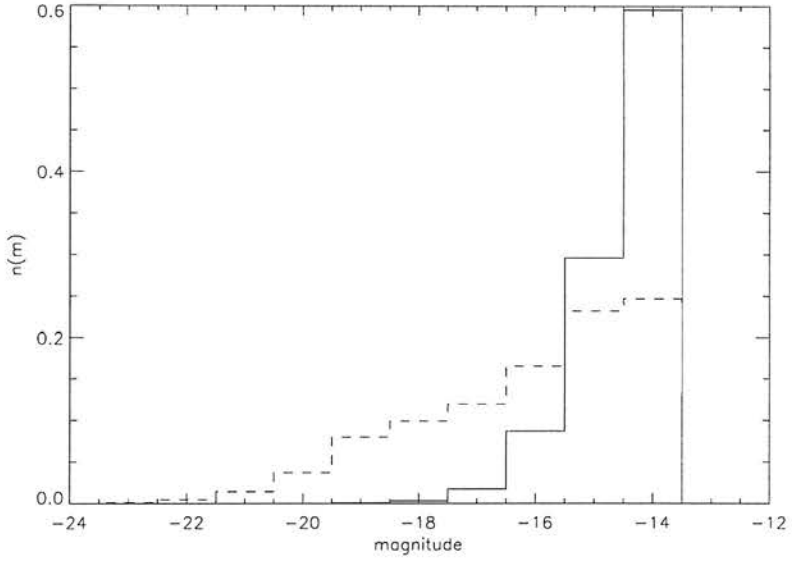

Figure 4.36: A comparison of the normalised (to 1) background magnitude histograms for stellar objects (solid line) and galaxies (dashed line) within the HIPASS survey region.

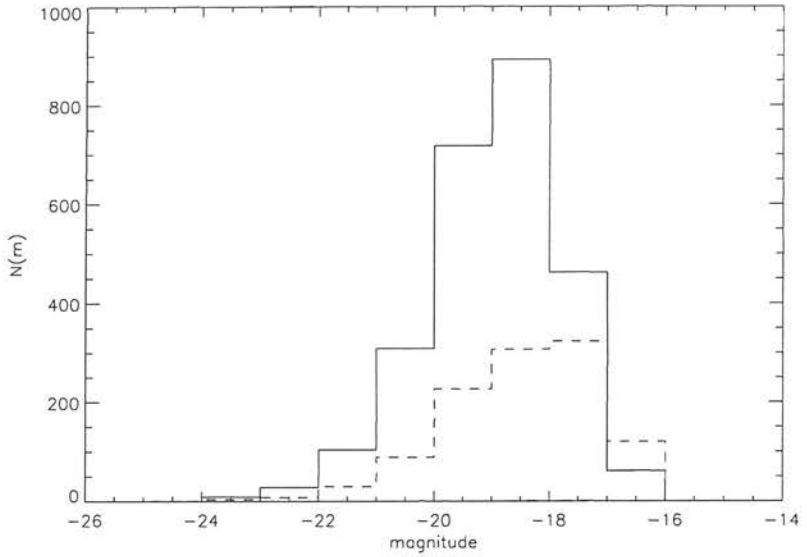

Figure 4.37: A comparison of the magnitudes of the stellar (dashed line) and galaxy (solid line) HIPASS source IDs selected by the EM algorithm.

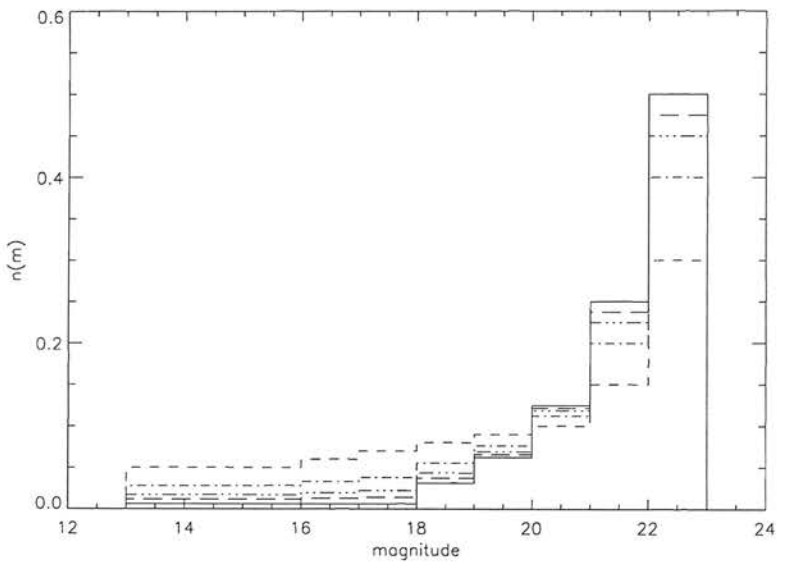

Figure 4.38: Model background magnitude histograms. Model A - solid line, model B - long dashed line, model C - dash dot dot dot line, model D - dash dot line, model E - short dashed line.

Table 4.4: Performance of likelihood ratio and EM algorithm on model data. Number in brackets after the EM algorithm performance is the number of sources to which it assigned IDs.

	% true IDs identified	
Model	Likelihood ratio	EM algorithm
A	97	95(103)
B	94	92(103)
C	86	84(101)
D	87	0(0)
E	77	0(0)

method. It can then be applied to the remainder of the source catalogue to find counterparts for unmatched sources. However, it is not a usual situation to have a partially matched catalogue which therefore makes this method not widely applicable. The SVM generated 1209 new matches bringing the total number to 3096 out of 4315, which is approximately 70 per cent.

The unsupervised EM algorithm was able to find optical IDs for 86 per cent of 4312 HIPASS radio sources using $\omega = 1$ magnitude, without the use of training data. It also selected 96 per cent of the velocity confirmed HOPCAT IDs and agreed with 95 per cent of the SVM IDs. We therefore infer that the spectral information used to compile HOPCAT was not necessary for discovering the optical IDs for these sources since the EM algorithm was able to identify them and also to identify counterparts for many other sources without any such information. However, the algorithm did select a number of stellar IDs for the radio sources, due to the vast proportion of stars in the field, which are possibly mis–identifications since it was given no star/galaxy information unlike the SVM. This presents an interesting avenue for future work to incorporate star/galaxy classification as a learned parameter into the EM algorithm.

The EM algorithm was also used on the HIPASS data in a supervised manner through use of the training data, i.e. explicitly telling the algorithm which IDs were already known, for direct comparison with the SVM method. With a magnitude bin width of $\omega = 1$ magnitude a performance equal to that of the SVM was obtained, as IDs were made for 70 per cent of the catalogue. In this trial the EM algorithm also selected far fewer stellar IDs (591, of which 304 were velocity confirmed HOPCAT IDs).

The HIPASS catalogue was randomly split into smaller subcatalogues from 80 per

cent to 3 per cent of its original size. The scatter in the percentage of IDs returned by the algorithm increased as the catalogue size was reduced, therefore, the results are less robust as the sample size decreases. However, errors on the learned magnitude probability distributions of the IDs calculated from these trials were small even in the 3 per cent catalogue case.

There was an average of 1.8 objects per source with $z_i^r > z_0^r$, i.e. which were more likely to be IDs for the source than the source in question not have a counterpart in the optical catalogue. For the 1893 sources with more than one object in their error ellipse with $z_i^r > z_0^r$, the ratio z_{ID}^r/z_2^r, the z_i^r value of the next best ID, were examined. 39 per cent of these were found to have $z_{ID}^r/z_2^r > 2$. Where there are two objects with very similar responsibility values there may be multiple IDs and both objects must be considered when using the results of such an association procedure.

The likelihood ratio made associations for 58 per cent of the HIPASS sources with $P_{ran} < 0.2$ of which the EM algorithm IDs agreed with 95 per cent. Where the two methods chose different IDs the likelihood ratio tended to select slightly brighter objects which were further from the source than the EM algorithm IDs. This was due to the learned magnitude probability distribution being relatively flat over the magnitude range -19.5 to -17.5, therefore, the EM algorithm distinguished between objects based on their distance from the source.

The likelihood ratio P_{ran} values were compared to the EM algorithms' z_{ID}^r. Where the two methods selected the same ID there was a clustering of objects with low P_{ran} and large z_{ID}^r indicating that these were reliable IDs in both methods. Where the methods selected different IDs there was a spread in both P_{ran} and z_{ID}^r. Out of the 109 IDs over which these methods disagreed in 100 cases the likelihood ratio ID had $z_{ID}^r > z_0^r$ but was not the chosen EM algorithm ID. The likelihood ratio also had difficulty selecting an ID for this small number of sources; the majority of the P_{ran} values of their IDs was very similar to that of the second best object and for 78 of these the second best ID also had a P_{ran} less than the cut–off value of 0.2.

We also note that by calculating z_0 the EM algorithm explicitly gives the cut–off value for good associations which had to be rather arbitrarily chosen in the likelihood ratio method. This is well illustrated in Figure 4.26 where we see that there are no IDs with $z_i^r < 0.1$. The selection of the P_{ran} cut–off can be difficult where the data is noisy as was the case for the FIRBACK survey application.

The EM algorithm did not perform as well on the small FIRBACK source catalogue making IDs for all 103 sources of which only 46 per cent agreed with the likelihood

ratio IDs. The discrepancy in performance of the algorithm on the two datasets was due to the shape of the magnitude histograms and therefore the value of the ratio $h_t/P(m_i^r|z_i^r = 0)$. The magnitude distribution of the SSS data in which associations were sought for the HIPASS sources had a long tail to bright magnitudes due to the properties of the field. At bright magnitudes, therefore, $h_t/P(m_i^r|z_i^r = 0)$ was large allowing selection of bright IDs. In contrast the INT–WFS data for the FIRBACK field had an overall much flatter distribution. The values of $h_t/P(m_i^r|z_i^r = 0)$ were therefore much smaller, hindering the choice of bright IDs.

For the FIRBACK sources there was an average of 40 objects per source with $z_i^r > z_0^r$, the algorithm was therefore having difficulty selecting an ID and the z_i^r values of the IDs were relatively low since there were several probable associations. The ratio z_{ID}^r/z_2^r was examined and it was found that only 30 per cent of sources had $z_{ID}^r/z_2^r > 2$. The values of P_{ran} and z_{ID}^r for the likelihood ratio and EM algorithm agreeing and disagreeing IDs were compared. In both cases there was a spread in P_{ran} and z_{ID}^r. For the 39 cases where the two methods disagreed the P_{ran} values of the likelihood ratio ID and second best match were compared. These were found to be very similar and 18 sources had P_{ran2} below the cut–off of 0.4. Therefore, there was more than one probable counterpart for these sources which the two methods were differentiating between.

4.8 Conclusions

In conclusion, the EM algorithm is a very effective association tool and in the case of the HIPASS survey it was able to select more IDs for the radio population than both the Mann et al. (1997, 2002) likelihood ratio and SVM methods. It also has two major advantages over the likelihood ratio method; firstly no assumption for the shape of the source optical counterpart magnitude distribution is necessary and secondly the algorithm calculates for each source a probability that there is no optical counterpart, therefore, explicitly giving the cut–off responsibility value for a good association. The EM algorithm is also more widely applicable than the SVM method since there is no requirement for any previous associations to have been made for use as training data. However, the FIRBACK survey was a particularly challenging association problem and the EM algorithm was not as effective on the FIRBACK survey data. This was due to the shape of the INT–WFS background magnitude distribution caused by the field selection. The likelihood ratio method was also not able to select IDs for a number of

sources in this case. There were several sources with more than one probable counterpart which the two association methods had to distinguish between. However, the EM algorithm method is more sensitive to the shape of the background magnitude histogram than the likelihood ratio method.

Both the FIRBACK and HIPASS datasets posed unusually challenging association problems due to their large positional errors, leading to numerous (up to several hundred) counterpart candidates to choose between within their error ellipses. They provided prime examples of the kind of situation where an automated statistical approach to finding IDs must be followed. Both the likelihood ratio and EM algorithm methods have shown that they can fulfil this need when used in the correct circumstances.

To err is human but to really foul things up requires a computer.

Anonymous

5

Implementing Association Algorithms On AstroGrid

5.1 Introduction

As described in Chapter 1 and illustrated in Chapter 3, to understand the properties of an astronomical source a model must be constructed which can explain its emission at all wavelengths. Astronomers are therefore increasingly interested in bringing together observations at many wavelengths. Due to the enormous expansion in survey astronomy in recent years, data at many wavelengths are available through survey archives. The concept of the Virtual Observatory is to bring together these archives so they can be accessed through a uniform interface such that, although held at distributed locations, they appear as one. This will enable users to interrogate multiple data centres in a seamless and transparent way and facilitate the multiwavelength astronomy which the astronomical community is striving for (AstroGrid Proposal 2001).

However, as discussed in Chapter 1, all the efforts of global VO projects will be in vain if it is not possible to associate entries in different databases, as the extra knowledge is based on comparing characteristics of a single object from different databases at

different wavelengths. The VO therefore requires a tool to make associations. During the work for previous chapters such tools were developed as IDL applications that ran on the desktop on local data. A similar facility was required for the VO. We therefore built an association tool and tested it on AstroGrid, the UK's VO project (see Chapter 1).

5.2 AstroGrid Architecture

Figure 5.1 shows the main components of AstroGrid in a UML (Unified Modelling Language) diagram, each component being represented by a rectangle with two further rectangles on one side. There are three client applications through which a user may access to AstroGrid. The Portal is a web based user interface, accessed through a web browser. The first AstroGrid Portal is hosted at Leicester but others are being set up at other locations such as Edinburgh. The Workbench or Astro Client Runtime (ACR) is a Java–based desktop application. This requires the user to install software on their desktop, which is not necessary when using the web portal. However, since it is Java–based it provides a more pleasant user experience. The third client application is the command line interface (CLI) which behaves like a UNIX terminal. To use this the ACR workbench must be installed anyway, as the ACR runs python scripts to get the results of command line requests through the workbench.

As mentioned in Chapter 1 Registries are essentially the 'yellow pages' of AstroGrid, listing all available resources. These are XML–based and conform to IVOA standards. They 'harvest' each other copying each other's contents, so all resources are listed in every registry. Registered resources have an associated key called an International Virtual Observatory Real Name (IVORN) by which they can be identified e.g. `ivo://org.astrogrid/alphabet#abc`. This is a form of URN (Uniform Resource Name) which are globally unique names for resources or units of information independent of their location. In the IVORN above the prefix `ivo:` indicates that this is an IVORN. The second part, `org.astrogrid`, is known as the 'authority ID' and is a unique name for the organisation which published the resource to the registry. The third section, `alphabet`, is known as the 'resource ID' which is the name of the resource and is unique within the publishing authority ID. The authority ID/resource ID combination are therefore a globally unique key to a particular resource: other authorities publish resources with the same resource name but they will not be confused since the authority IDs will be different. Each authority ID is managed by one registry but

that registry may manage several authority IDs (VOTech DS6 Tool Integration Report 2005[1]). Anything after the # is not part of the IVORN, but is extra information that can be understood by the resource itself. The registry associates an IVORN with an XML entry that describes the resource to which it corresponds. This is called the VOResource and contains certain minimum metadata such as the contact details of the person who entered it, a human–readable name and description etc. There are also a number of standard extensions to the schema describing certain types of resource. For example, an application VOResource will contain the input and output parameter names and data types (DS6 Tool Integration Report 2005).

The Community component is involved in user authentication (establishing if the user is who they claim to be) and authorisation (establishing what they have permission to do). There are lots of communities which can be managed locally e.g. there is one at Leicester, one at Edinburgh, one at Cambridge etc. Each user is a member of a

[1]available at http://wiki.eurovotech.org/bin/view/VOTech/Ds6ToolIntegrationReport

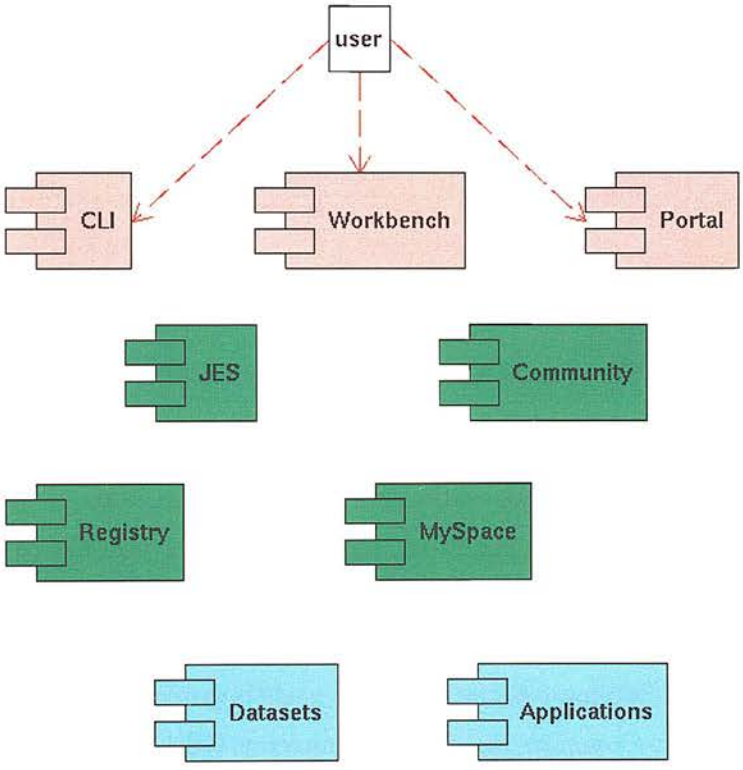

Figure 5.1: AstroGrid component diagram.

community. This can be used to restrict access to certain resources to authorised communities. There will be an IVORN representing the community which manages that user. Figure 5.2 shows a UML sequence diagram of the login process. Time progresses down the page and actions are shown in chronological order, a rectangle stretching from one action to a returned result represents a process that must be completed before another may be begun. The user logs in to the web portal using their username, password and community. The Portal then contacts the Registry and requests the Registry entry for the user's specified Community name and the VOResource for that Community service is returned. The Portal extracts from this the URL of that Community service. It then contacts the Community service with the user's login name and password. If the community service recognises these it returns a security token which can then be passed around between other AstroGrid components when the user is submitting jobs to prove their authentication.

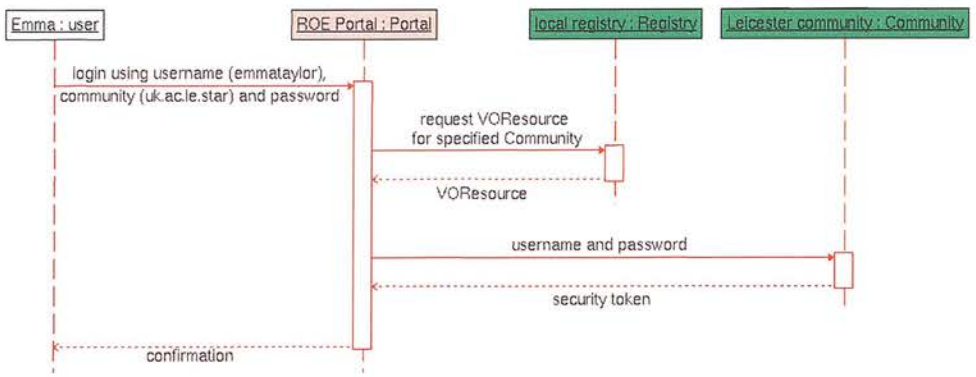

Figure 5.2: Sequence diagram for the Portal login process.

AstroGrid is essentially a cooperating set of web services each of which is entered into the Registry (DS6 Tool Integration Report 2005). A webservice is 'a software system designed to support interoperable machine–to–machine interaction over a network. It has an interface described in a machine processable format.'[2]. Therefore to make applications available on AstroGrid they must be wrapped as webservices. This is achieved through a Common Execution Connector (CEC) which is a special type of webservice that provides extra interoperability required by AstroGrid. A CEC is an instance of a CEA (Common Execution Architecture) which is a model of how applications are run in the VO, including a definition of the tool, its parameters and

[2]W3C definition available at: http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/

description of how to execute it.

Applications are registered with the Registry in two ways. The first describes the application in a non–location–specific way and contains the application metadata (i.e. the methods available, the number of parameters and their types) and the second is an entry for the webservice that provides the application and will contain the URL for the CEC (i.e. the application's location). An application may be installed at several locations e.g. the ROE, Cambridge or Leicester, and therefore be listed in several CEAService registry entries.

Figure 5.3 shows a UML representation of some applications. The Edinburgh CEC has two applications for which it provides a CEA interface (represented by a circle) in order to make them available to the grid. The Cambridge CEC also provides one of these applications. Thus there is one application registry entry for the 'sextractor' application, and two service entries that include reference to the sextractor application entry. Thus we know that one application is available from two services.

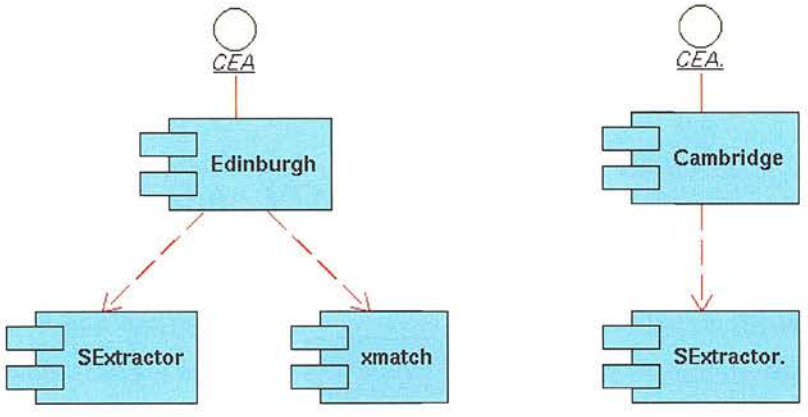

Figure 5.3: Component diagram for AstroGrid applications.

There are many different types of databases that are linked together by AstroGrid (e.g. MySQL, MS SQL Server, PostgreSQL) which all speak slightly different dialects of the database query language SQL. An extra component is therefore necessary to translate user queries into the correct dialect. This is known as the Data Set Access (DSA) component and has a variety of interfaces through which other components may interact with it. Figure 5.4 shows a UML component diagram for a database; here it has five example interfaces.

There is a CEA interface and a SkyNode interface. These provide similar access and

are both being developed into IVOA standards, at present the CEA is implemented by AstroGrid and SkyNode by the American VO. The SIAP (Simple Image Access Protocol) is for publishing and accessing image data. The Cone Search interface provides simple radial searches to be made around a specified position. The browser interface provides direct access through a web browser (DS6 Tool Integration Report 2005).

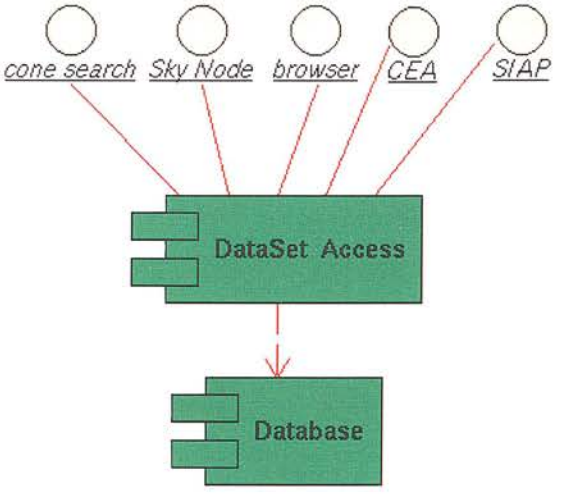

Figure 5.4: Component diagram for AstroGrid databases.

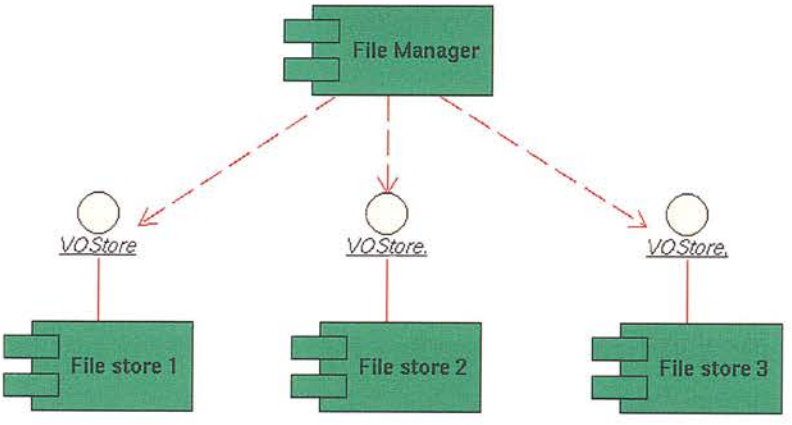

Figure 5.5: Component diagram for MySpace.

The Job Execution Service (JES) manages jobs being executed by the AstroGrid

system and provides the user with status information about progress. A job is submitted to JES as a workflow document, and JES calls the various CEA services in the requested order. When a job is submitted which uses an application the JES sends a request to the Registry which returns a list of the locations of all CEAServices which offer that application. It can then decide which location to submit the job to, considering factors such as load balance and network 'nearness' of the application. Network 'nearness' describes the speed of moving information from A to B and is related to the network speed, it is not necessarily related to physical proximity. JES also allows the user to manipulate results using a sophisticated scripting engine based on Groovy (DS6 Tool Integration Report 2005).

The virtual storage space, MySpace, enables users to store workflows, queries and results remotely rather than having to download them to the desktop. Figure 5.5 shows a UML component diagram of MySpace. To the user MySpace appears as any standard file manager with all their files stored in one place. The file manager in fact only stores metadata about the files. In reality their files may be distributed over several different file stores which each have a standardised VOStore interface (DS6 Tool Integration Report 2005) to take advantages of the 'nearness' of the stores to point of use.

5.3 Review of other VO Association Work

The Euro–VO VO–Tech project has recognised the need for a cross matching tool in the VO. It has a number of design study areas one of which, DS4, is focussing on user tools. Under this study a set of requirements for a VO cross matching tool have been drawn up (Requirements for a Cross Matching Tool[3]) and simple matching algorithms discussed.

The DS4 group has focussed on spatial matching suggesting a simple nearest neighbour approach. Building on this a slightly more evolved method could include treatment of positional errors. However, we believe spatial matching is not adequate in the general case since it does not take into account how common or rare the chosen object may be. The DS4 study has also suggested providing the user with the ability to specify some customised 'distance' function which may include extra information such as magnitude, type etc. This would provide a simple means of weighting distance but is still not entirely satisfactory.

It is pointed out in the study that likelihood estimates for associations are very

[3]http://wiki.eurovotech.org/bin/view/VOTech/CrossMatchReq

useful scientifically allowing users to seek more reliability. However, calculation of likelihoods will reduce performance since extra information will have to be calculated e.g. local source density. In such a case the user could also provide their own likelihood function. It is clear to us that the increased reliability of associations derived from a likelihood method far outweighs any concerns about reduced performance. Astronomers are familiar with computationally expensive procedures and are used to having to wait for results, we do not think that a short delay in receiving cross associated IDs would discourage astronomers from using a cross matching tool. Providing users with the ability to specify their own likelihood function would increase usage flexibility. However, providing a number of tried and tested tools such as our own will speed up research and prevent astronomers from having to design their own algorithms which is an area they may be unfamiliar with.

There has also been work in the USA on a cross matching service built into SkyQuery[4]. This is a prototype of a federated database application which uses the Sloan Digital Sky Survey (SDSS, Abazajian et al. 2005), 2MASS (Jarrett et al. 2000) and the FIRST radio survey (Becker et al. 1995). These databases are held at different geographical locations and all implement the American VO SkyNode interface. The SkyQuery system consists of a 'Client' which is a simple web page used for user access and a 'Portal' which executes the distributed queries by splitting them up between the individual SkyNodes.

At the heart of the SkyQuery distributed query engine is a spatial cross matching algorithm. This has been incorporated into an SQL procedure that can be called in an SQL query that searches several databases. SkyQuery was instrumental in establishing the IVOA SkyNode standard and it is a requirement that all SkyNodes are able to perform this spatial cross matching algorithm.

Figure 5.6 shows an overview of the SkyQuery architecture. In this system each database indicates its willingness to participate in query processing by registering itself with the Portal. The user is then able to submit a query via a Client to the Portal. The portal parses the query into sub–queries which are specific to each database. These sub–queries are first transformed into queries which count the number of rows that the query will produce. This count brings the data into the cache. When the full query is executed it will be hitting the cache and therefore be much faster, thus minimising network traffic. The database returns the query results back as XML packaged as a DataSet. The cross matching function takes the resulting DataSet from each SkyNode

[4]www.skyquery.net

queried, writes it to a temporary table and performs the algorithm.

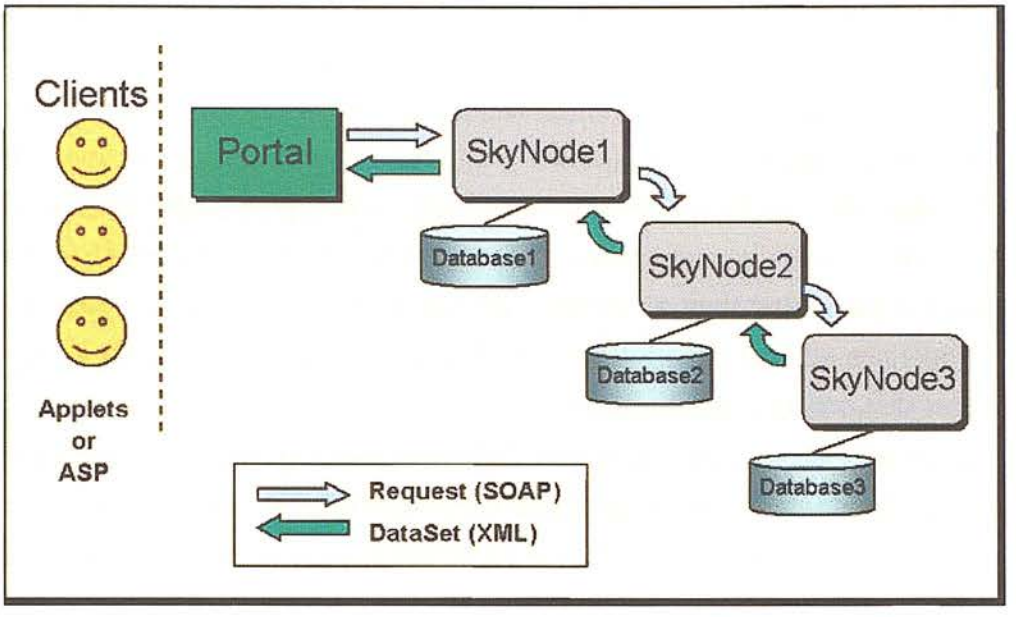

Figure 5.6: Overview of the SkyQuery architecture. Clients connect to the portal which executes the distributed query on the database SkyNodes and receives XML DataSets. (Diagram from www.skyquery.net)

This method of doing the cross associations at the database locations reduces the need for the movement of large amounts of data which makes it very efficient. However, this is still a purely spatial matching method and if other factors were incorporated to make it more widely applicable further computation would be required.

Another approach is exemplified by the SSA, available through AstroGrid, which also provides CrossNeighbours tables[5]. These are tables recording the ID numbers of sources in other databases lying within 10 arcsec of each SSA entry. There are tables providing nearby objects in the 2MASS point source and extended source catalogues, the SDSS DR1 and EDR catalogues and the USNOB catalogue. These could be used as a first stage in an association procedure as the search for objects within a certain distance of the SSA detections has already been done. A more computationally expensive cross association algorithm could then be used on these tables to identify the IDs.

Work has also been carried out at CSIRO in Australia to do fast catalogue matching based on spatial proximity, the results of which are then intended to be subject to further tests based on further data such as redshift. With modest computing equipment

[5]http://surveys.roe.ac.uk/ssa/www/ssa_browser.html

they claim to be able to spatially match two catalogues containing a billion objects in less than 6 hours. Their technique is based on plane sweep and filter–and–refine methods to search through the catalogues for possible matches (Abel et al. 2004).

The plane sweep technique sorts the catalogues by declination and then searches for objects in a catalogue C that match a specific object from a catalogue B using the filter–and–refine method. The filter step finds answers to a weakened form of the problem that can rapidly generate a set of candidates which is guaranteed to include all pairs of objects (hits) that satisfy the full form of the problem but also some objects (false drops) not satisfying the full criteria. The refinement step then selects only the candidates that satisfy the full criteria.

In this catalogue association case the full form of the matching problem is to report all pairs of objects whose angular separation is less than some threshold i.e. objects from catalogue C are reported that are contained within a circle with a radius determined by the object positional errors centred on an object in catalogue B. The weakened form of the problem is to report all objects from catalogue C contained within the bounding box of the object from catalogue B's circle defined by the limits of the circle right ascension and declination (Abel et al. 2004).

Their algorithm employs a plane sweep technique to identify the candidate pairs by maintaining an 'active list'. This is a list of objects from catalogue C that might match the current object from catalogue B or further objects from B, and this list is maintained to ensure that the search is executed efficiently. The size of the active list is restricted via a sequential traversal of C.

A basic outline of the algorithm is given below:

1. Pre–process the catalogues B and C, sorting into ascending order by declination;

2. For each object from B,

   Delete from the active list any members that cannot match the current member of B or any further members of B, on the basis of declination;

   Insert into the active list any members of C that might match the current object from B (b) on the basis of declination;

   Search within the active list for matches against b, this is a search on right ascension;

3. Terminate by flushing any remaining members of the active list (Abel et al. 2004).

In summary, starting at the bottom of the sorted catalogue B, an active list is set up which contains all objects from C with declination within b's bounding box (this is the filter step). The active list is then searched for objects which are situated within b's error circle (this is the refine step). This results in a list of counterpart pairs for b. For the next object in B (b') objects from C with declination less than b's upper limit just have to be added to the top of the active list and those which are below its lower limit deleted from the bottom. This is a very efficient system since a new active list does not have to be created with each new object from B that is considered and it is only the small active list that is considered for matches not the full catalogue C.

This algorithm produces a list of counterpart pairs. The results of this full catalogue matching process would be very useful in large scale work as the first stage in an association process. The results could then be checked using further criteria.

Table 5.1: UCDs for VOTable contents.

UCD	source table	candidate table	magnitude histogram table	results table
ID_MAIN	source names			
UCD for RA e.g. POS_EQ_RA_MAIN	source RAs	counterpart RAs		association RAs
UCD for Dec e.g. POS_EQ_REC_MAIN	source Decs	counterpart Decs		association Decs
UCD for magnitude e.g. PHOT_MAG_B		counterpart magnitudes	magnitudes of histogram bins	association magnitudes
ID_PARENT		corresponding source names		corresponding source names
UCD for name e.g. ID_NUMBER		counterpart names		association names
STAT_PROBABILITY				figure of merit for matches
NUMBER			number in magnitude bin	

Table 5.2: Parameters required by the three different association methods

Parameter	Poisson	Likelihood Ratio	EM Algorithm
source list	√	√	√
candidate list	√	√	√
random source list		√	
candidate list for random sources		√	
magnitude histogram	√	√	√
survey area used to produce magnitude histogram/unit area	√		
completeness limit of counterpart catalogue	√	√	√
search radius	√		√
σ for Gaussian distribution of source positional uncertainty		√	√
initial estimate of γ			√
initial estimate of counterpart magnitude histogram			√
RA UCD	√	√	√
Dec UCD	√	√	√
magnitude UCD	√	√	√
counterpart name UCD	√	√	√
return number	√	√	√

5.4 Development of an Association Tool for AstroGrid

We developed an association tool offering three different association methods: the Mann et al. (1997, 2002) likelihood ratio method (see Chapter 2); the Poisson method (see Section 1.7); and the EM algorithm (see Chapter 4). These three applications are known as xmatchLikelihoodRatio, xmatchPoisson and xmatchEM respectively.

As described earlier, data are passed around in a standard format in the VO, known as VOTable, which is an XML mark–up standard for astronomical tables. The data within these VOTables are described by Unified Content Descriptors (UCDs) which form a formal vocabulary for astronomical data controlled by the IVOA. Using UCDs avoids the proliferation of terms, synonyms and reduces ambiguity (IVOA VO Architecture Overview 2004[6]). They are also designed to be understood by both humans and computers. The xmatch applications were designed to work with data in VOTable format and identify the information they require within the tables using the UCDs. The source list, candidate list, magnitude histogram and results table are all parameters required or returned by the association applications in VOTable format. The UCDs

[6]available at http://www.ivoa.net/Documents/latest/IVOArch.html

describing the information held within these are given in Table 5.1, these clearly show how the UCD describes the quantity it is associated with: for example PHOT_MAG_B describes a photometric magnitude in the B-band. An example VOTable is presented in Figure 5.7. This contains the information about two sources listing their RAs, Decs and names, each data field or 'column' has an associated datatype, name and UCD.

The three xmatch applications - xmatchLikelihoodRatio, xmatchPoisson and xmatchEM - were written in Java for ease of installation on AstroGrid. To handle the VOTable data we used the Starlink Tables Infrastructure Library (STIL[7]) which is a set of Java class libraries which allow input, manipulation and output of such tabular data. In the Poisson association method it is possible to calculate the required *a priori*

[7]http://www.star.bris.ac.uk/~mbt/stil/

```
<?xml version='1.0'?>
<VOTABLE version="1.1"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://www.ivoa.net/xml/VOTable/v1.1 http://www.ivoa.net/xml/VOTable/v1.1"
 xmlns="http://www.ivoa.net/xml/VOTable/v1.1">
<RESOURCE>
<TABLE name="VOTable_example" nrows="2">
<FIELD datatype="double" name="RA" ucd="POS_EQ_RA_MAIN"/>
<FIELD datatype="double" name="DEC" ucd="POS_EQ_DEC_MAIN"/>
<FIELD arraysize="26" datatype="char" name="NAME" ucd="ID_MAIN"/>
<DATA>
<TABLEDATA>
  <TR>
    <TD>283.49292</TD>
    <TD>-78.8525</TD>
    <TD>HIPASSJ185358-785109_08415</TD>
  </TR>
  <TR>
    <TD>325.49583</TD>
    <TD>-17.471944</TD>
    <TD>HIPASSJ214159-172819_05817</TD>
  </TR>
</TABLEDATA>
</DATA>
</TABLE>
</RESOURCE>
</VOTABLE>
```

Figure 5.7: An example VOTable. This includes the information for two sources listing their RAs, Decs and names using appropriate UCDs to describe the data.

probability using the data for one source at a time. However, the EM algorithm requires the data about all sources at once for its iteration procedure and, although it is possible to calculate the P_{ran} values in the likelihood ratio method for one source at a time, this requires a list of random sources for comparison. The applications therefore require the source candidate lists as one long concatenated table containing the information about all possible counterparts for the entire source list rather than separate lists for each source. Hence the source candidate VOTables must be concatenated and an extra column added containing the name of the source for which the objects are a candidate for identification purposes.

The final list of parameters required by each method is given in Table 5.2. These must be provided as primitive data types (e.g. doubles, Strings etc.) on AstroGrid, therefore any files are read in as one long String. In each association application the user is also required to specify how many possible matches they wish to have returned for each source (n). The results are returned as a concatenated VOTable which contains information about n matches for each source, sorted such that the most likely match appears first. Each counterpart has an associated figure of merit for the match stored with UCD STAT_PROBABILITY. The meaning of this figure will depend on the method being used, it will either be a P_{ran}, an *a priori* probability or a responsibility value if the user has selected the likelihood ratio method, Poisson method or EM algorithm respectively.

The Java classes written for each of the three applications were given a public interface to provide the point of entry into each application. This prevents users accessing the internal code. One overall class was constructed for all three, `XMatchAstroGridInterface`, which contained a method for each application: `xmatchPoisson`, `xmatchLikelihoodRatio` and `xmatchEM`. The required xmatch project Java classes were compressed into a single jar file for installation on AstroGrid using the CEA–installer. This wrapped the applications into a webservice with a CEC and also registered them with the Registry. The CEC uses Java–reflection to introspect the overall interface class extracting meta–information to generate the registry entry for the application. Java–reflection is not able to deduce parameter names, only method names, number of parameters and parameter types. It was therefore necessary to edit the XML registry entry manually for each of the xmatch applications to give the parameters human readable names and descriptions.

5.5 Portal Tour and Application Testing

The login screen a user is presented with when accessing AstroGrid through the web portal is shown in Figure 5.8. A username, password and community must be entered here to login. Once these are verified through the process described earlier and illustrated in Figure 5.2 the user reaches the home screen, shown in Figure 5.9. This gives a list of current jobs and provides access to other areas of the portal.

Through the workflows screen it is possible to construct a workflow containing steps, loops etc. Basic workflows were constructed for each of the xmatch applications to test them on the AstroGrid system. Once a step is added to a workflow the user can browse the registry for a resource they wish to use e.g. one of the xmatch applications. A parameter box then appears listing the parameters required for the application, the names and datatypes of these are found by the portal in the VOResource document supplied by the Registry.

The simple workflow used to initially test the xmatchEM application is shown as an example in Figures 5.10. The three applications were tested on selecting IDs for a small subset of 100 HIPASS sources from a list of candidates from SuperCOSMOS. VOTables for these data and a SuperCOSMOS magnitude histogram were constructed and uploaded into MySpace. This allowed direct comparison with results achieved previously for this data with IDL code (see Chapters 2 and 4).

Parameters for applications may be 'direct' i.e. a value directly entered into the workflow parameter box, or 'indirect' i.e. a reference to a file stored in MySpace. It is possible to browse MySpace from the workflow to find files to be used as parameters. MySpace may also be browsed directly through the portal using the MySpace screen, shown in Figure 5.11. It is possible to upload and download files to and from the desktop here and also upload files directly from the web.

Once a workflow has been constructed it can be submitted to AstroGrid for execution by JES and its progress monitored via the job monitor screen, shown in Figure 5.12. This receives constant updates of the job status from JES which manages the job. Once the job is complete the results may be found in MySpace. It is also possible to view a workflow transcript listing the stage–by–stage progress of the job in detail which is useful when investigating problems.

The three xmatch applications returned their resulting source associations to MySpace in VOTable format. These results were exactly the same as those produced by the original IDL applications.

A sequence diagram for one of the basic one–step workflows is shown in Figure 5.13. This shows how the workflow is submitted to JES from the Portal, and JES then requests the Registry entry for the location of the application required. Once JES has decided on which location to send the job to, by extracting the location from the VOResource document supplied by the Registry, it sends it to the application in the form of a 'tool document'. This is an XML document which is embedded in the workflow and contains the values of the parameters and any other information needed to run the application. The application then requests any files it needs from MySpace, performs its task, returns the results to MySpace and notifies JES on completion. Meanwhile the Portal regularly requests status information from JES and updates the user on the progress of their job.

The web portal also provides a data query builder, shown in Figure 5.14. This helps the user to construct queries for available datasets in ADQL, the astronomical data query language (IVOA Astronomical Data Query Language Version 1.01[8]). The example shown is a query requesting five columns of information (ID, RA, Dec, B-magnitude and a blend flag) about objects within a given radius of a given source position. Queries can be saved to MySpace and then added as a step in a workflow the results of which will be returned as a VOTable.

[8]available at: http://www.ivoa.net/Documents/latest/ADQL.html

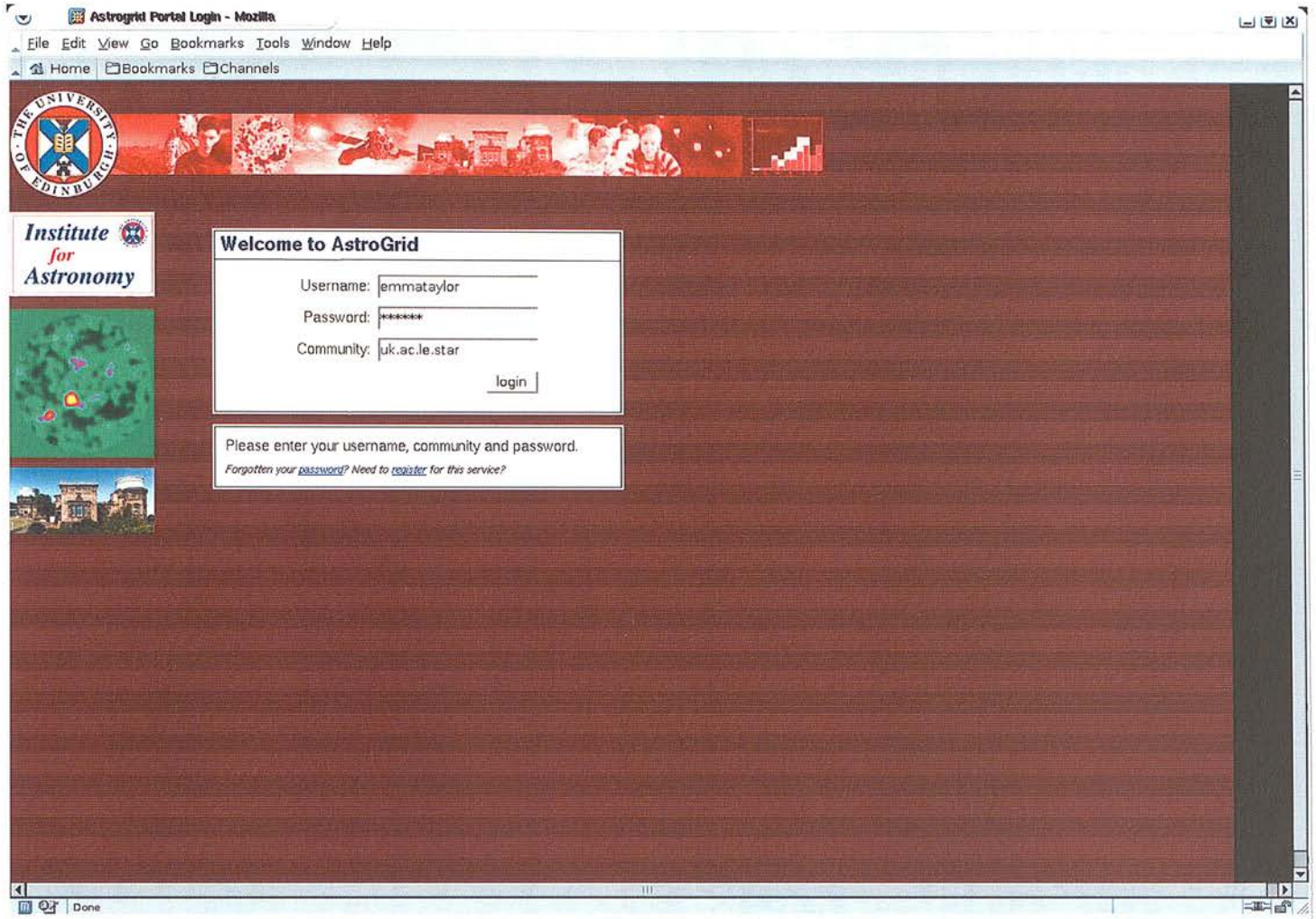

Figure 5.8: The ROE AstroGrid portal login screen.

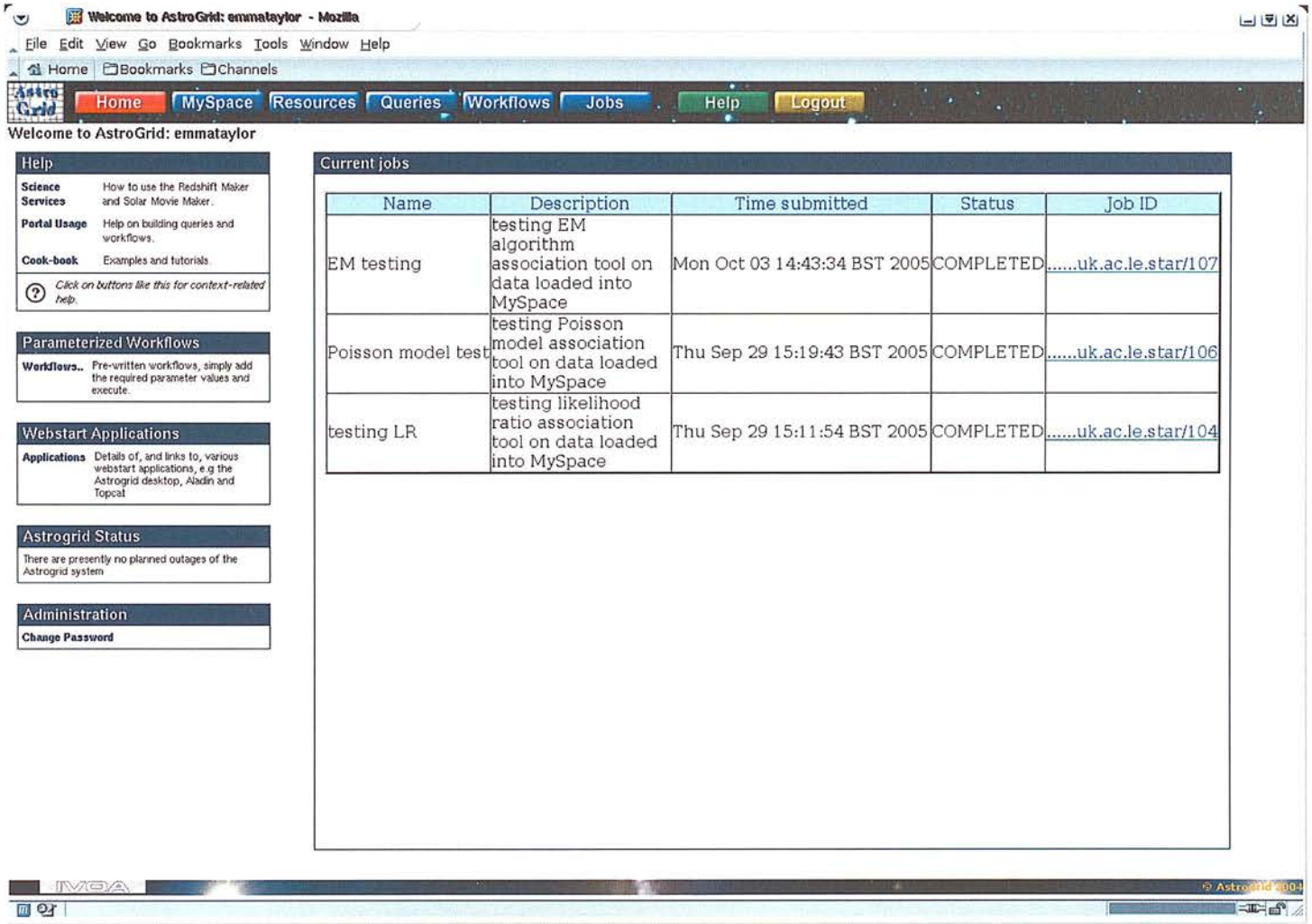

Figure 5.9: The AstroGrid portal home screen.

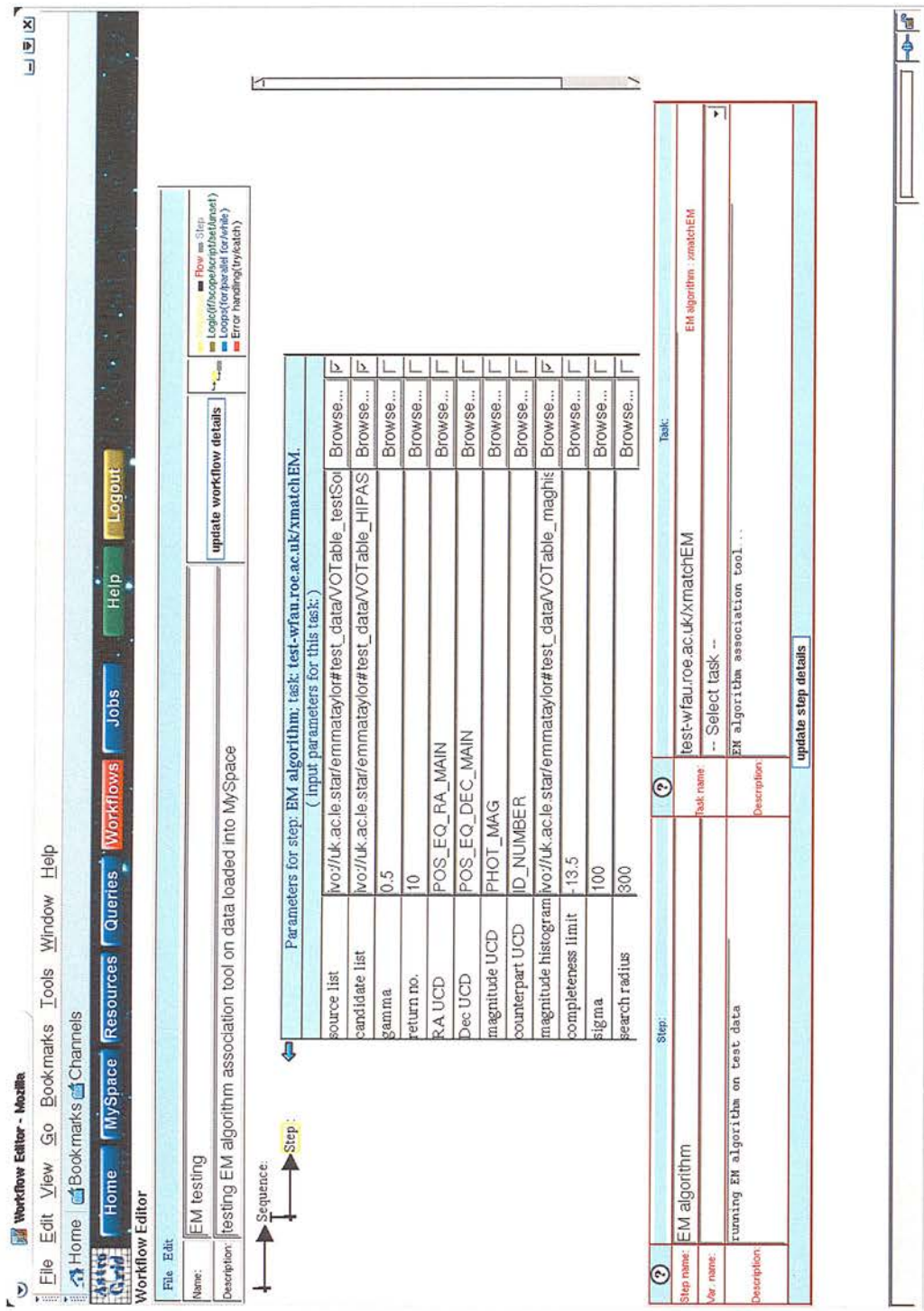

Figure 5.10: A workflow used to test the EM algorithm association tool.

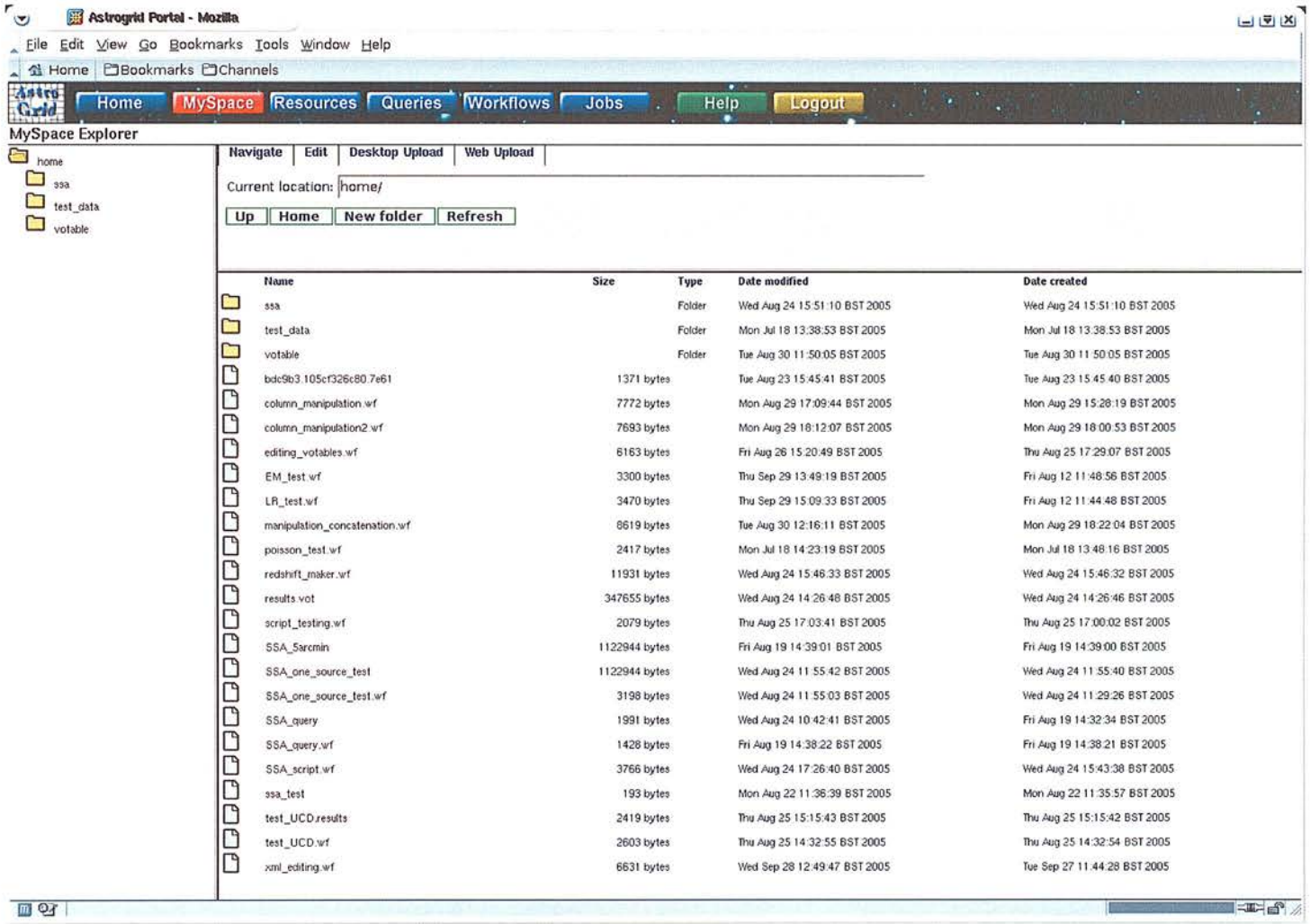

Figure 5.11: The AstroGrid MySpace screen.

Figure 5.12: The AstroGrid job monitor screen.

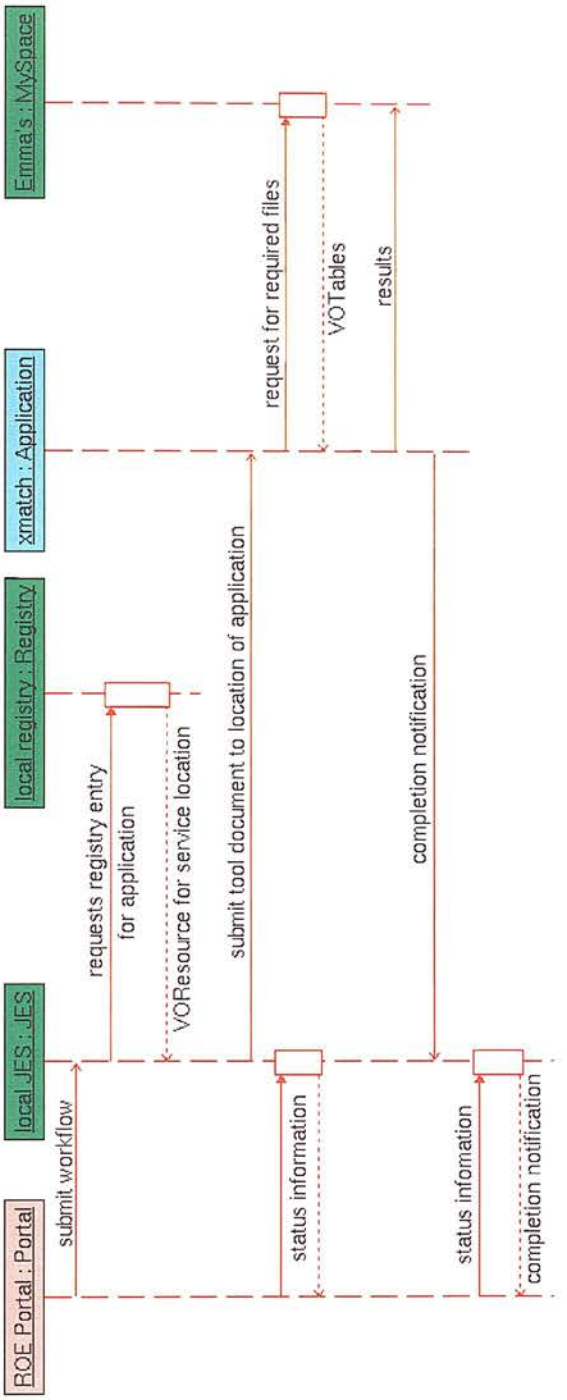

Figure 5.13: A sequence diagram for a simple one step workflow used for testing the xmatch applications.

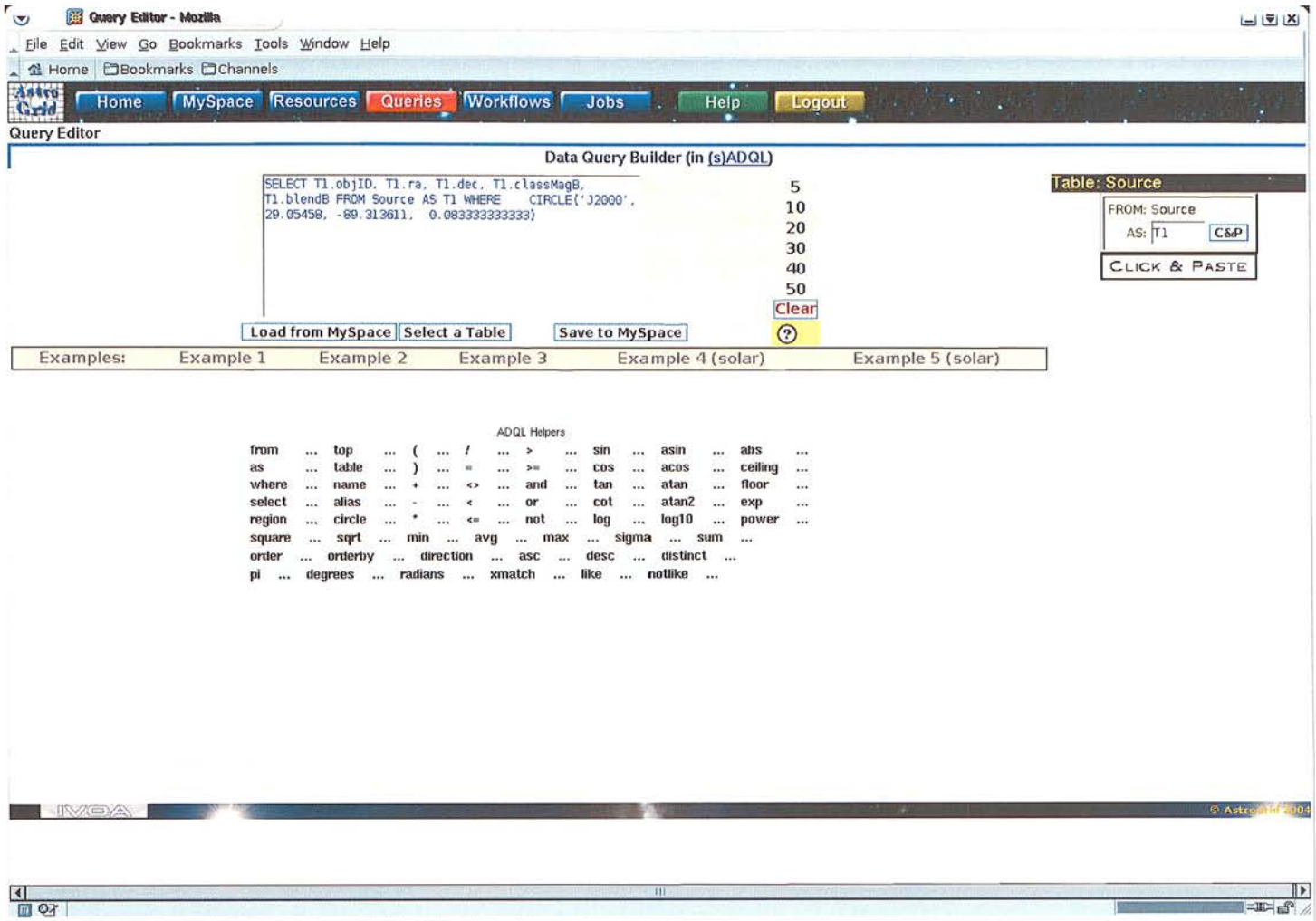

Figure 5.14: An ADQL query constructed through the AstroGrid Portal.

5.6 Using the xmatch Tool on AstroGrid

Typically an AstroGrid user would use one of the xmatch applications in a situation where they have a list of sources for which they wish to find associations in a dataset available on the system. In such a case there are several pieces of information that must be extracted from the relevant database: the completeness limit of the observations; the magnitude distribution of the dataset; candidate lists for the sources; and, for the Poisson association method, the survey area. At AstroGrid's present stage of development acquiring this information is not trivial. Several ADQL commands are currently not supported in the Portal or are malfunctioning, namely MAX, MIN and CIRCLE. Therefore, for our work the magnitude distribution was constructed by hand through the database's direct browser interface which allows use of MAX and MIN.

Since the CIRCLE ADQL command was malfunctioning we queried the required database directly using an http get command. This is a way of encoding a request into a URL which then points to the results. There is a limit imposed by each data centre on the number of lines returned by an ADQL. These are intended to prevent people accidentally downloading the whole database and are set quite low. These limits are inhibitive when the search radius around a source position is large such as the 5 arcmin for the HIPASS sources and had to be increased for our work.

The dataset completeness limits and survey areas are, at present, unavailable through the AstroGrid system and must be found in the literature. However, this information will eventually be included in the dataset metadata, the form of which is currently being worked on by the IVOA. Once the magnitude histogram for a dataset has been calculated it would also be computationally economic to store it for future use rather than every user recalculating it.

To execute a task such as that outlined above a workflow must be constructed that does the following:

1. queries the relevant database requesting the required information about candidates within a certain radius of each source

2. resulting candidate lists must be concatenated into one long candidate list and an extra column added to indicate which source each entry is a candidate for

3. concatenated candidate list and other parameters given to an xmatch application to find IDs for the sources

We tested the three applications in this realistic situation by searching for IDs for

100 HIPASS radio sources in the SuperCOSMOS Science Archives (SSA) database that is available through AstroGrid. An example of the workflow constructed to do this is shown in Figure 5.15.

A workflow is constructed as a sequence of tasks set out in order down the screen beginning at the *Sequence* arrow. In the example the first four tasks are *Set* commands which set up arrays containing the RAs, Decs and names of the 100 sources and also a constant value for the radius within which associations were sought (5 arcmin in this case). The *For* loop then looped over each of the 100 sources. Within the loop there was a *Sequence* of two tasks which were carried out for each source. The *Script* formed the URL for downloading the catalogue of objects within the set search radius of the RA and Dec of the source in question. It read in all the information about these candidates which comprised of 57 columns of data. The *Script* also made cuts for magnitude, star/galaxy class and a blend flag. This removed objects for which magnitudes were not able to be calculated and those for which star/galaxy class was unclassifiable or classed as noise. When SuperCOSMOS plates are first scanned objects very close to each other may be detected as one large object known as a parent object. After deblending these may be split into separate objects, known as children. The blend flag cut ensured that only non–deblended objects and children were included in the final catalogues. These cuts reduced the number of rows in each table from over 1000 to a few 100. The resulting tables were saved to MySpace.

Once the individual source candidate lists had been constructed the next *Step* within the loop extracted the required columns from each table and added the column containing the source name indicating for which source the object is a candidate. This reduced the number of columns from 57 to just 5: RA, Dec, magnitude, candidate name and source name. A Java application called `xmatchQueryResultsManipulation` was written to do this since it was not possible through scripting; is now listed in the Registry. As parameters this application requires: the candidate list file, the name of the corresponding source, the UCDs for RA, Dec, magnitude and candidate name. It uses the UCDs to select the required columns from the large table and returns a VOTable file containing the five necessary columns to MySpace.

After completing the loop and manipulating the candidate lists for the sources, a *Script* concatenated the tables to form the long candidate list required by the xmatch applications. The final *Step* was an xmatch association application.

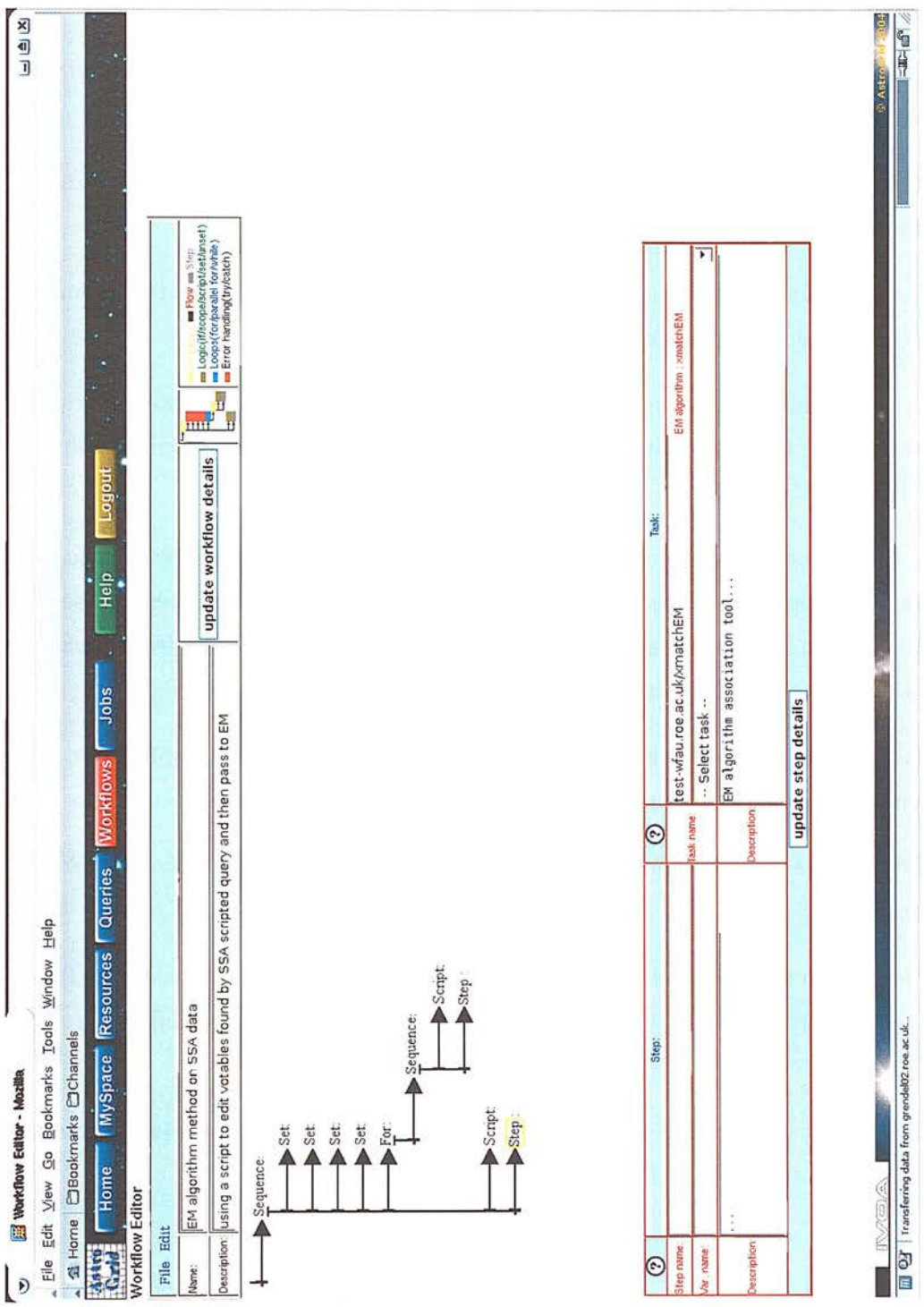

Figure 5.15: An example of a workflow used to find IDs for a list of sources from a database available through AstroGrid.

A sequence diagram for this workflow is shown in Figure 5.16.

- The workflow is submitted to JES which then performs a series of 'http get' requests to the SSA database.
- After each object catalogue is returned, JES executes the script to remove the excess rows and sends the resulting VOTable to MySpace.
- JES then requests the registry entry for the location of the xmatchQueryResults-Manipulation application.
- The URL for the service location is extracted from the VOResource document and the tool document sent to the application.
- The application requests the set of object catalogue VOTables from MySpace, executes its task and sends a VOTable that contains the four required columns of data to MySpace.
- At the end of the loop over the 100 test sources, once JES has received notification that the application has completed for the 100th time, it executes the second script. This involves requesting all the object catalogue VOTables from MySpace and constructing the long concatenated table.
- The final step in the workflow can then be executed, the xmatch application. JES requests the registry entry for the location of the xmatch service.
- JES submits the tool document to the location it extracts from the VOResource document.
- The xmatch application then requests the concatenated table and other VOTables it requires from MySpace, executes, sends the results to MySpace and notifies JES of completion.
- Meanwhile, as described earlier, the Portal regularly requests status information from JES and displays the progress of the users job.

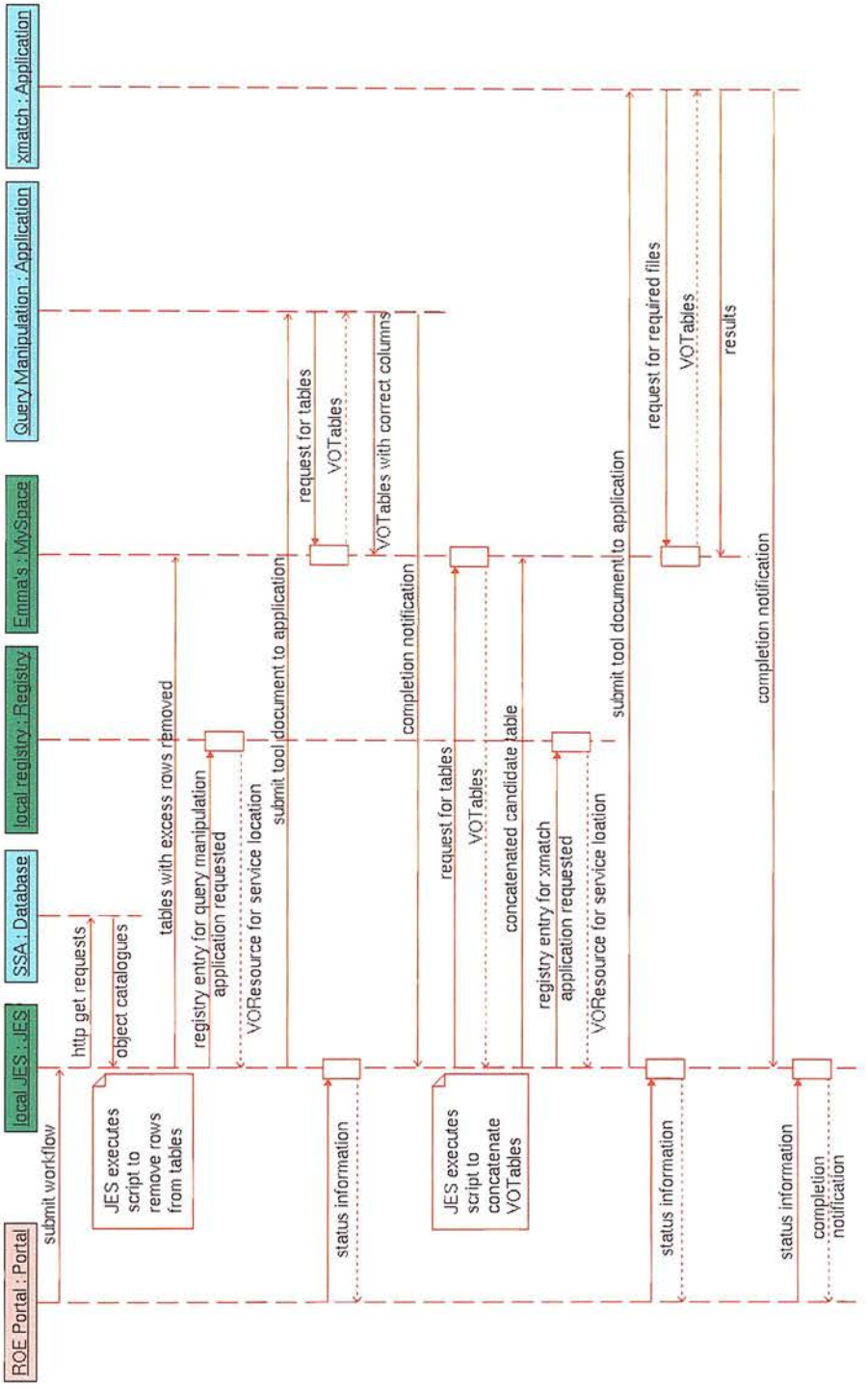

Figure 5.16: A sequence diagram for the workflow shown in Figure 5.15.

The SSA catalogues available through AstroGrid are object catalogues extracted from scans of photographic Schmidt survey plates by the SuperCOSMOS machine. The data used in the EM algorithm investigations in Chapter 4 were the results of SExtractor analysis of the images. The catalogues were therefore slightly different. For a direct comparison of the AstroGrid xmatch applications and the original IDL tools we used counterpart lists for the 100 HIPASS sources directly from the SSA[9]. The IDL tools were run on these catalogues and the IDs selected for the 100 HIPASS sources and their associated P_{ran}, *a priori* probabilities and responsibility values calculated by the three methods compared to those calculated by the xmatch tools on AstroGrid. The results were exactly the same.

The xmatch applications are clearly powerful tools that will be extremely useful to AstroGrid users. However, it is currently not straightforward to convert the results of database queries into the form required by the xmatch applications, as illustrated by the workflow described above. The table manipulation was complicated and computationally expensive, however it is preferable to reduce the size of the VOTables as much as possible before passing them to a Java application. The STIL library used to handle the VOTables is still in development and classes are being written to make the table manipulation easier and will be available shortly. There were also problems with script time–out as the time limits allowed for scripts to run on AstroGrid is only 10 minutes. This is to try to prevent users from doing heavy computation using JES. However, until the required ADQL commands are available this is unavoidable. For the xmatch application the time–out limit was increased.

5.7 AstroDAS

Making associations is often computationally expensive. It may therefore be economic to store resulting matches for re–use. Diego Prina Ricotti at the Università di Roma Tre has developed the Astronomical Distributed Annotation System (AstroDAS) which stores the results of cross association algorithms in a relational database as mapping annotations (Prina–Ricotti 2006). It also provides users with the ability to query the stored mapping annotations. Therefore, before embarking on the lengthy development and runtime of a cross–associating workflow users can search those that have already been done.

Consider the galaxy NGC7108 as an illustration. This is a large nearby

[9]SSA radial search: http://surveys.roe.ac.uk/ssa/radial.html

Table 5.3: A mapping table example.

SDSSDR2	USNOB	2MASS
587727214959198470	3573413503478	1331337280
587726879952667039	3577708427512	1331337184
587727214959198583	3573413503561	1331337393
587726879952732671	null	1331355633
587726879952666873	3577708427409	null

galaxy which has been detected by many different surveys at many wavelengths and therefore has an entry in many different astronomical survey databases. For example, in the SDSS data release 2 database it has ID 587727214959198470 which may be represented as the database object $\langle sdssdr2, 587727214959198470\rangle$. In the USNOB sky survey database the same galaxy has an entry $\langle usnob, 3573413503478\rangle$. These two entries, therefore, 'map' to each other: this is denoted $\langle sdssdr2, 587727214959198470\rangle \rho \langle usnob, 3573413503478\rangle$. A mapping table can be constructed to summarise such relations, as shown in Table 5.3 (Prina-Ricotti 2006). The first line of this is equivalent to:

$$\begin{aligned} \langle sdssdr2, 587727214959198470\rangle \quad \rho \quad & \langle usnob, 3573413503478\rangle \\ \langle usnob, 3573413503478\rangle \quad \rho \quad & \langle 2mass, 1331337280\rangle \\ \langle sdssdr2, 587727214959198470\rangle \quad \rho \quad & \langle 2mass, 1331337280\rangle \end{aligned}$$

We note that these mapping relations are also reflexive (an object is in a mapping relation with itself i.e. $o_1 \rho o_1$), symmetric (the order a relation is expressed in is unimportant i.e. $o_1 \rho o_2 = o_2 \rho o_1$) and transitive (if $o_1 \rho o_2$ and $o_2 \rho o_3$ then $o_1 \rho o_3$).

The AstroDAS system is based on a distributed peer–to–peer architecture with each node of the network known as an 'annotation peer'. Each annotation peer has its own database for storing cross–match results and a web service interface that provides methods for both storing associations in, and querying, the underlying database. This structure allows different institutions, groups or even individuals to maintain their own database containing their own mapping annotations. Each peer can be registered in a VO Registry just like any other database thereby making its annotations available to VO users.

Along with the mapping annotations the database stores the author of the annotations or the name of the algorithm which produced them, the time they were entered into the database, the reliability of the match and the mean RA and Dec values of the

two detections in question. It is therefore possible for AstroDAS users to query the system for matching objects published only by algorithms or individuals they trust and with reliabilities above a specified threshold.

A vector of mapping annotations can be stored in an annotation peer database using the `insertMappingAnnotations` method. An example of this process is illustrated in Figure 5.17. Each AstroDAS installation consists of three components: a client, a portal/peer and a database. When attempting to store mapping annotations in the database, the client component interprets the user's `insertMappingAnnotations` request and sends an 'http post' message to the annotation peer. The body of the http document contains a SOAP (Simple Object Access Protocol) message within which are held the mapping annotations in XML format. The AstroDAS peer then queries the relevant astronomical databases for the positions of the objects, in order to calculate the mean RAs and Decs of the object pairs. It then sends the mapping annotations and extra data to the underlying database to be stored. If the process is successful a confirmation message is returned to the client otherwise an error message is sent.

A user may query all the entries in the AstroDAS peer–to–peer network using SQL. All these entries, from the local mapping tables of all the nodes, form what is known as the 'global mapping table'.

Annotations available on the network may also be used to query the external astronomical databases using a form of SQL developed specifically for this purpose, Distributed SQL (DSQL). This allows users to specify whose annotations they trust or set a threshold on the reliability of the annotations which they accept. An illustration of such a query is given in Figure 5.18. The AstroDAS client component sends a web service request to the portal; which is a web service interface for executing DSQL queries. The portal interprets the message and generates a query for the peer–to–peer network, resulting in a mapping table between the databases specified in the DSQL query. The mapping table is then used by the portal to query the astronomical databases using the object ID numbers from the table to identify the required objects in the astronomical databases. When the data are received by the portal it joins the tables from the astronomical databases using the mapping table and sends the results back to the client. This method of querying the astronomical databases using the object ID numbers relies on these numbers being fixed. We note that although this is sufficient most of the time, since most astronomical databases are read–only, for databases which have incremental updates or data releases/amendments this could cause confusion. It would therefore be useful to include in the AstroDAS database schema some version informa-

tion and likewise in the VOResource metadata describing the astronomical database in the Registry.

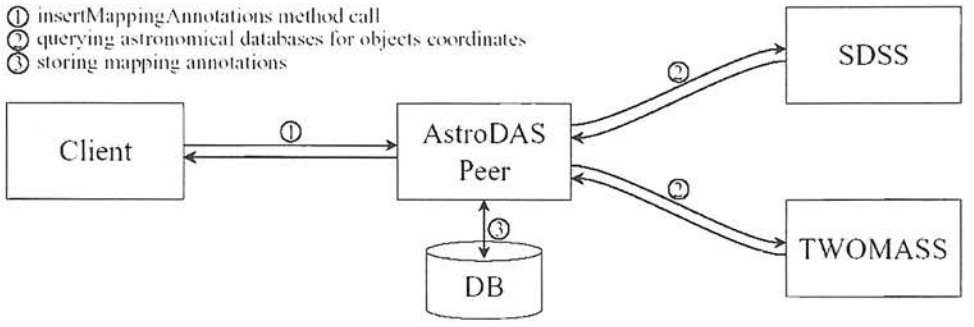

Figure 5.17: An illustration of how mapping annotations are stored using AstroDAS (from Prina–Ricotti 2006).

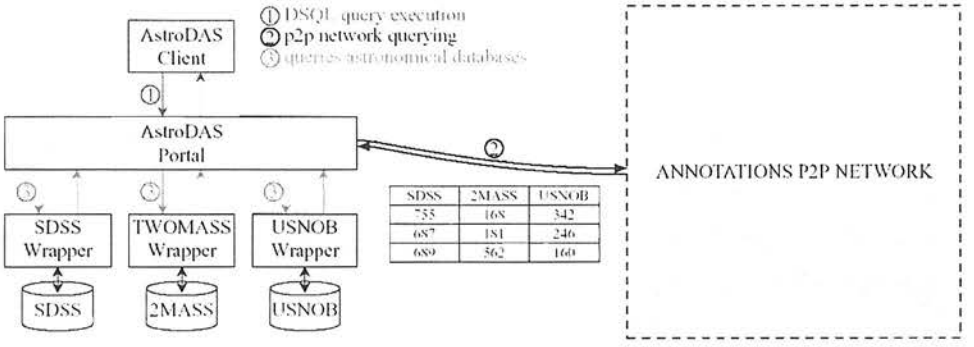

Figure 5.18: A schematic diagram of a DSQL query being processed by AstroDAS (from Prina–Ricotti 2006)

We set up our own annotation peer to enable us to store the results of the HIPASS/SSA cross matching that had been executed via AstroGrid. A new Java application, `AnnotationsPopulator`, was developed to convert the xmatch application results into the format required by AstroDAS and use the `insertMappingAnnotations` method to populate our annotation peer. This was deployed on AstroGrid to allow users to incorporate it into a workflow as the final step in a sequence which cross–associates two datasets, storing the results for future use. It takes as parameters the xmatch application's resulting VOTable, the UCDs for RA, Dec, magnitude and counterpart name, the names of the two cross–matched databases and the author of the associations e.g. `xmatchEM`, `xmatchLikelihoodRatio` or `xmatchPoisson`.

However, since the HIPASS sources were not stored in a database we adapted the AstroDAS software slightly to prevent it attempting to query the astronomical databases when annotations are inserted, in order to calculate the mean RA and Dec. We note that in the majority of cases mean RA and Dec values will be of little use since the spatial resolution of the two surveys in question will be very different. In the HIPASS/SSA matching case the optical SSA positions are much more meaningful since the positional errors on the HIPASS sources are large ($\sim$ 0.8 arcmin). We therefore altered the `insertMappingAnnotations` method such that the RA and Dec values to be stored may be specified by the user, only if they are not given are the astronomical databases queried for object positions. Our situation, where a user has a source list with which they seek associations from a database, is a common one and in such case a database for the sources is not available.

Once our HIPASS/SSA cross–associated objects had been inserted into the annotation peer database we were able to query it directly using SQL. Since object positions, matching reliabilities and the author of the annotations i.e. the algorithm which produced them had been stored, it was possible to use these as criteria in a query.

Using DSQL to query the SSA astronomical database using criteria from the annotation database was non-trivial. AstroDAS relies on the astronomical databases having a SkyNode interface. At present the SSA has a SkyNode interface with only partial functionality which could not be used in this case. We therefore had to adapt the AstroDAS software to use the CEA interface which is fully functional on the SSA. Figure 5.19 shows an example of a DSQL query submitted through the AstroDAS web client. It asks for the object ID, RA and Dec of objects from the SSA database within a certain radius of a specified position, which are also stored in the Edinburgh University annotation peer and have been cross associated using `xmatchEM`. Figure 5.20 shows the results of such a query.

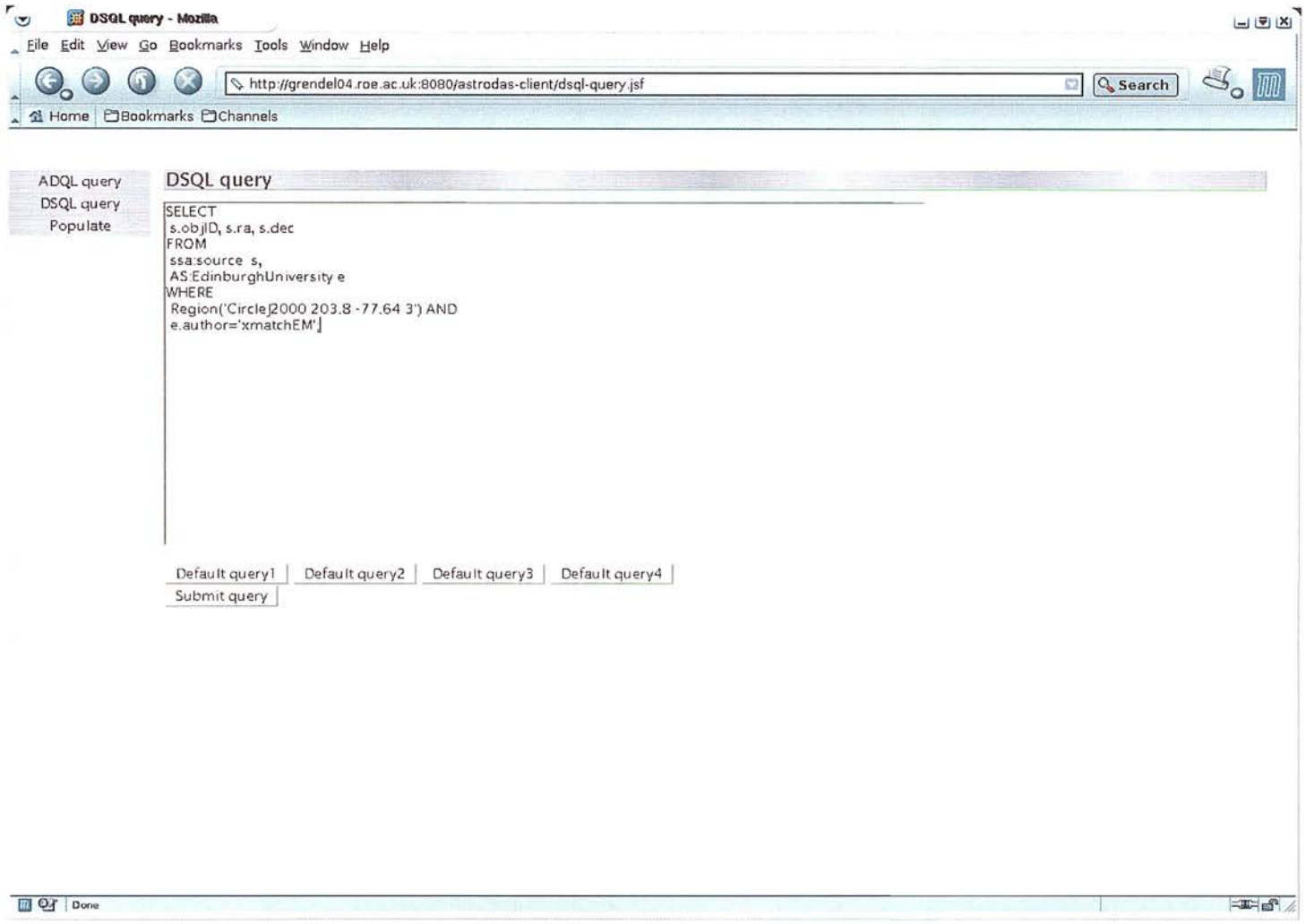

Figure 5.19: Submitting DDQL query to AstroDAS system via the AstroDAS web client.

Figure 5.20: Results from DDQL query.

5.8 Conclusions

We have developed an association tool for AstroGrid offering three different methods: the likelihood ratio method, the EM algorithm and the Poisson method. These are known as `xmatchLikelihoodRatio`, `xmatchEM` and `xmatchPoisson` respectively. Such services are vital to the global VO initiative since without the ability to determine which object in one catalogue is associated with that in another catalogue the potential knowledge available from multiwavelength the characteristics of objects will not be extractable. The value of multiwavelength work in astronomy is becoming more and more apparent and therefore these tools and others like them will be heavily relied on.

The three xmatch tools were tested thoroughly on finding optical IDs in SuperCOSMOS data for 100 HIPASS sources on both data uploaded onto AstroGrid and that from the grid–based SSA service. The results were compared to those found by the previously used IDL tools and found to be exactly the same. The xmatch tools may now, therefore, be made available to the general astronomical community on AstroGrid, empowering users to cross–associate the numerous datasets available and thereby extract new science.

However, due to the current developmental stage of the AstroGrid project, using the association tools is not entirely straightforward. The applications require several input parameters regarding the data which are not readily available. The dataset completeness limits and survey areas must be found in the literature. Due to ADQL problems the magnitude distribution must also be constructed by hand rather than via an AstroGrid workflow and the queries made using an http get request. This resulted in a lot of complicated scripting and another Java application having to be written to manipulate the query results into the form required by the xmatch tools. In the near future, however, these problems will be resolved. There is work underway to add to the STIL library used to manipulate VOTables which will simplify the scripting, and the number of ADQL commands supported is being increased.

The IVOA is developing a standard for resource metadata at present. For example, to describe a survey there will be:

- identity metadata which includes the survey name, curation metadata such as the name of the publisher and their contact details;

- general content metadata which contains a general description of the resource, collection;

- service content metadata which includes information about the survey coverage and resolution;

- data quality metadata which describes general uncertainties in the data;

- service metadata which give information about any services offered by the dataset such as a URL interface (Resource Metadata for the Virtual Observatory Version 1.1 2005[10]).

The data quality metadata includes a *spatial uncertainty* entry which could be taken as σ for the Gaussian distribution of positional offsets in the xmatch applications. The content metadata also includes a *depth of coverage* entry, however, this may not be the same as the survey completeness limit and the *spatial coverage* entry does not contain a simple piece of information for calculating the survey area. We would therefore recommend that the IVOA standard include the completeness limit and survey area of a catalogue. It would also be useful to include a recommendation for the counterpart search radius based on the positional errors of the dataset contents. These metadata are typically not queryable parameters and even though they will become included in the data description they must be discovered through other means by a VO user (Resource Metadata for the Virtual Observatory 2005). We would advocate providing this information directly through the VO. Once the magnitude histogram for a dataset has been calculated it would also be computationally economic to store it for future use rather than every user recalculating it. A similar method to AstroDAS of attaching such information to data would be extremely useful on AstroGrid as it could save a lot of CPU time.

Cross–matching algorithms all require similar parameters such as the background catalogue magnitude histogram. It would therefore be sensible to set out a basic standard structure for VO cross–matching applications. We recommend that these follow the structure of the three xmatch applications requiring the source list, concatenated candidate list and magnitude histogram as VOTables and offering the user the ability to specify the number of matches to return for each source. The results VOTable should contain the probability of the match as calculated by the method in question sorted such that the most likely match appears first.

Finally, the three xmatch applications need associated documents to describe the different methods and recommend under which circumstances to use each method. We

[10]available at http://www.ivoa.net/Documents/latest/RM.html

suggest that the Poisson method be used when choosing between only a few candidates for each source since it is a simple model. The likelihood ratio method is preferable when the survey area is restricted therefore producing a background magnitude histogram with a restricted shape. The EM algorithm is very effective when the survey area is large and the population for which IDs are sought is unknown. Documentation can be supplied in a number of different ways on AstroGrid. There is a wiki on which information can be posted and a user guide is also being worked on. The other option for an application such as this is to wrap it as a parameterised workflow. This provides a workflow a user has already been constructed to which they can simply alter the parameter values. However, since AstroGrid is changing so rapidly documentation will be added once it is more stable.

We implemented the Astronomical Distributed Annotation System (AstroDAS, Prina–Ricotti 2006) which stores the results of cross–association algorithms as mapping annotations in a database. Therefore, before embarking on the lengthy development and runtime of a cross–associating workflow users can search those that have already been done. The system is based on a distributed peer–to–peer architecture with each node of the network known as an 'annotation peer'. This allows users to publish and manage their own sets of cross–matched objects. We set up our own annotation peer and published to it the results of the HIPASS/SSA cross–matching using an application, `AnnotationsPopulator`, developed to convert the results of the three xmatch applications into the form required by AstroDAS and populate the annotation peer database. This application has been deployed on AstroGrid and is published in the registry allowing users to store their own cross–match results in their own annotation peer, it can be incorporated as the final step in a cross–matching workflow.

You Lord, in the beginning created the earth, and with your own hands you made the heavens. They will disappear, but you will remain; they will all wear out like clothes.

Hebrews 1 v.10-11

6 Conclusions

6.1 The Need for an Association Tool in the Virtual Observatory

In recent years astronomers have faced an ever increasing mountain of data which is the result of four converging trends: 1. the vast amounts of data produced by new instruments driven by Moore's law; 2. huge amounts of data from simulations; 3. the ability to store petabites of data on–line; 4. the internet and Grid which makes these archives accessible to anyone anywhere (Gray and Szalay 2002). Astronomers are therefore already accustomed to doing research on–line through such data centres, however, it is no longer practical to download a copy due to the size of the datasets available.

The UK Virtual Observatory project, AstroGrid, aims to accelerate the quality of on–line research. Data centres will provide further services beyond simply downloading subsets of their data, enabling users to make complex queries. It will also be possible to analyse data in situ to avoid the movement of vast amounts of data. These improved services are part of a long term trend in astronomy to develop standard tools which free the astronomer from mundane tasks such as writing their own reduction code so

they can concentrate on science (AstroGrid Phase A Report 2002)[1].

With this end in mind AstroGrid aims to link the archival datasets of space– and ground–based observatories and the catalogues of multiwavelength surveys and make all these databases interoperable so they are seamlessly, jointly browsable in a simple way. It will develop database analysis and exploration tools and also information discovery tools (AstroGrid Proposal 2001)[2] and provide the computational resources to run them. This will benefit the entire astronomical community from students and amateurs to professionals (Williams 2003).

Science progresses through the federation of information, bringing information from different sources into the same frame of reference (Williams 2003). This is certainly true in astronomy since to understand the astrophysical properties of a galaxy requires the construction of a model that can explain its emission at all wavelengths. This therefore requires the federation of observations from many passbands to disentangle the emission caused by different physical mechanisms (Mann and Williams 2004)[3]. Clearly the Virtual Observatory will empower astronomers to do this by allowing all the linked databases to be accessed through a uniform interface and although they are held at distributed locations they will appear as one. However, a key element required by the VO is the ability to associate entries in different databases to enable the multiwavlength data for celestial objects to be brought together. Without a method for doing this the enormous global effort to link so many datasets and provide tools for their analysis will be in vain.

This thesis has aimed to provide a solution to this problem by investigating several association techniques with a view to providing them in the form of a tool for AstroGrid users. The results of this research are summarised below.

6.2 The Likelihood Ratio

It is far from trivial to determine that radiation detected at one wavelength originates from the same object as that at another wavelength. This is especially difficult when the spatial resolutions of the two instruments are very different resulting in up to hundreds of possible counterparts for a source within its error ellipse. The likelihood ratio association method is a probabilistic approach able to use basic prior knowledge

[1] AstroGrid Phase A Report: http://wiki.astrogrid.org/pub/Astrogrid/PhaseAReport/redbook.pdf

[2] AstroGrid Proposal to PPARC 2/4/01: http://wiki.astrogrid.org/pub/Astrogrid/CoreDocs/astrogrid-proposal.doc

[3] 'What is the Virtual Observatory?': http://www.roe.ac.uk/~rgm/sc4devo/WhatIsTheVO.pdf

such as counterpart magnitudes to aid the association process. The likelihood ratio is given by:

$$LR = \frac{q(m,c)e(x,y)}{n(m,c)} \tag{6.1}$$

where q(m,c) is the magnitude and type probability distribution function for the true counterparts, e(x,y) is the distribution of their offsets from the source position and n(m,c) is the surface density of background objects with magnitude m and type c. This quantity is the ratio of the probability of finding the true ID at this position with this magnitude to the probability of finding a similar background object here by chance.

Two likelihood ratio source association methods were tested via searching for INT–WFS optical counterparts for the 175 μm FIRBACK source population; the Mann et al. (1997, 2002) prescription and the Rutledge et al. (2000) prescription. This was a challenging association problem due to an error of up to 65 arcsec on the source positions leading to hundreds of possible optical counterparts within each 175 μm source error ellipse.

The Mann et al. (1997, 2002) likelihood ratio prescription was found to be preferable in a case such as this with a small source sample size and where the optical images contain many extended objects from which an image analyser identifies numerous sources. In this method the likelihood ratio values of the candidate objects are compared to those for a large number of candidates around fictitious randomly placed sources to compute a probability, P_{ran}, for each candidate. P_{ran} is the probability that the likelihood ratio of the most likely association with a randomly placed source is at least as high as that for the true source. It is then necessary to choose a P_{ran} threshold value for confident associations rather arbitrarily from the sources P_{ran} histogram. In this particular application this selection was difficult as the histogram was very noisy due to the large source positional errors and the fact that the source list may contain spurious sources from the supplementary catalogue. Where positional errors are smaller and the existence of sources more confident the selection of the P_{ran} threshold can be made easily.

The Rutledge et al. (2000) likelihood ratio prescription may be applied when associating over 1000 sources with counterparts from a different catalogue. In this method the probability that an object is the unique association for the source, $P_{id,i}$, for each possible counterpart is calculated as well as P_{no-id}, the probability that the source has no counterpart in the catalogue. This, therefore, gives a quantified threshold for the source having no association which may be preferable to that arbitrarily chosen in the Mann et al. (1997, 2002) method.

6.3 The Nature of the 175 μm FIRBACK Sources

Having selected the appropriate assocation method, we proceeded to apply it to the FIRBACK catalogue to identify them with optical IDs from the INT–WFS catalogue using the Mann et al. (1997, 2002) likelihood ratio technique, described above. To understand the nature of celestial objects it is necessary to examine their emission over a wide range of wavelengths. It is then possible to compare their spectral energy distributions (SEDs) to those predicted by models to gain an insight into their nature. The properties deduced can have a significant impact on theories of the evolution of individual galaxies and the universe as a whole. After identifying them with optical IDs we investigated the nature of the 175 μm FIRBACK sources. This was led by myself in the N2 field (Taylor et al. 2005) and Michel Dennefeld in the N1 field (Dennefeld et al. 2005).

In the N2 field we found optical identifications for 33 out of 55 FIRBACK sources. Associations were then also sought with sources in the multiwavelength ELAIS band-merged catalogue at 6.7 μm, 15 μm and 90 μm with associated data at U,g',r',i',Z,J,H,K and 20 cm (Rowan-Robinson et al. 2004). 60 μm and 100 μm fluxes or upper limits were also obtained for each of the sources via the IRAS SCANPI facility[4]. This yielded a final set of 31 associations not only with optical IDs but also with detections at up to nine other wavelengths.

The nature of this population was investigated through a comparison of their observed spectral energy distributions (SEDs) with predictions from radiative transfer models (Efstathiou and Rowan-Robinson 2003) which simulate the emission from both cirrus and starburst components. We found 80 per cent of the far–infrared sources to be star–bursting galaxies with their starburst component at a high optical depth. The resulting SEDs were also used to estimate far–infrared luminosities, star formation rates, dust temperatures and dust masses. The N2 FIRBACK population was found to consist of four suspected ultra–luminous infrared galaxies (ULIRGs) with $L_{FIR} \sim 10^{12}$ $L_{\odot}$ and $SFR_{FIR} > 100$ $M_{\odot}yr^{-1}$, a number of luminous infrared galaxies (LIRGs) with moderate star formation rates and $L_{FIR} \sim 10^{11}$ $L_{\odot}$ and a population of low redshift quiescently star forming galaxies.

The estimated dust temperatures of the starburst galaxies are 30-49 K which is relatively low compared to that of standard ULIRGs which have $T_D \sim 50$ K. The most luminous of these galaxies also have large dust masses which supports galaxy

[4]http://irsa.ipac.caltech.edu/applications/IRAS/Scanpi/

evolution models in which these cold, dusty, luminous objects have large masses of dust heated at a moderate intensity (Chapman et al. 2002). The mix of higher redshift ULIRGs and lower redshift starburst galaxies is also in good qualitative agreement with the predictions of the Lagache et al. (2003) model starburst templates as adopted by Sajina et al. (2003) who investigated a small sample of N1 FIRBACK sources. Lagache et al. (2003) also state that to gain agreement between their models and observations they require a cold local population which is what we have found.

The properties of these galaxies are similar to those of the SCUBA population, offering a stepping stone between high redshift SCUBA galaxies and the local universe. The cold dust temperatures of the FIRBACK sources may seriously affect the interpretation of the SCUBA sources. Since many models rapidly evolve a hot ULIRG population to the local luminosity function they may predict a false redshift distribution if these cold sources are not considered.

6.4 The EM Algorithm

It is a common situation in astronomy to wish to investigate a source population for which we have no prior knowledge about the properties of the source counterparts expected at another wavelength, for example through observations with a new instrument. In such a case it is necessary to estimate the counterpart magnitude distribution to use the likelihood ratio association method. Since little was known about the FIRBACK sources prior to these investigations their optical magnitude distribution had to be estimated in order to assign them optical IDs. To alleviate this problem we developed a new astronomical application of a machine learning technique known as the EM algorithm which is used in the field of informatics. This is able to 'learn' the source magnitude distribution iteratively.

The algorithm was tested on the FIRBACK sources for direct comparison with the likelihood ratio method and also radio sources from the HI Parkes All–Sky Survey catalogue. This dataset had been used by a group at the University of Queensland to study (Rohde et al. 2005) other machine learning techniques including the Support Vector Machine (SVM) which they found to be most effective. We also applied the Mann et al. (1997, 2002) likelihood ratio method to these data to facilitate a comparison of the various association methods. The other advantage of this data set was that part of the HIPASS catalogue (HICAT) had already been identified with optical counterparts from the SuperCOSMOS Sky Survey through analysis of radial velocity data to produce

HOPCAT. This, therefore, gave us another means of testing if the likelihood ratio and the EM algorithm were giving the 'correct answers'.

The EM algorithm was a very effective association tool and in the case of the HIPASS survey it was able to select more IDs for the radio population than both the likelihood ratio and SVM methods. It also has two major advantages over the Mann et al. (1997, 2002) likelihood ratio method: firstly, no assumption for the shape of the source optical counterpart magnitude distribution is necessary; and, secondly, the algorithm calculates for each source a probability that there is no optical counterpart, therefore, explicitly giving the cut–off responsibility value for a good association. The EM algorithm is also more widely applicable than the SVM method since there is no requirement for any previous associations to have been made for use as training data. However, the FIRBACK survey was a particularly challenging association problem and the EM algorithm was not as effective on the FIRBACK survey data. This was due to the shape of the INT–WFS background magnitude distribution caused by the field selection. The likelihood ratio method was also not able to select IDs for a large number of sources in this case; both methods struggled since they found several objects that were probable counterparts for many of the sources but it would appear that the EM algorithm method is more sensitive to the shape of the background magnitude histogram than the likelihood ratio method.

Both the FIRBACK and HIPASS datasets posed unusually challenging association problems due to their large positional errors, leading to numerous (up to several hundred) counterpart candidates to choose between within their error ellipses. They provided prime examples of the kind of situation where an automated statistical approach to finding IDs must be followed.

6.5 An Association Tool for the VO

As described in Section 6.1 this thesis aimed to solve the problem of cross associating the large databases that are now common in astronomy. The research undertaken culminated in the development of an association tool for AstroGrid, offering three different association methods - the Poisson method, the (Mann et al. 1997, 2002) likelihood ratio method and the EM algorithm - known as `xmatchPoisson`, `xmatchLikelhoodRatio` and `xmatchEM` respectively. Such services are vital to the global VO initiative since without the ability to determine which object in one catalogue is associated with that in another catalogue the potential knowledge made available through the linking of vast numbers

of datasets will not be extractable. The value of multiwavelength work in astronomy is becoming more and more apparent as time goes on and therefore these tools and others like them will be heavily relied on in the future.

The three xmatch tools were tested thoroughly on finding optical IDs in SuperCOSMOS optical data for a subset of 100 HIPASS sources on both data uploaded onto AstroGrid and that extracted directly from the SuperCOSMOS Science Archive (SSA) database. The results were compared to those found by the IDL tools developed for the work in previous Chapters and found to be exactly the same. The xmatch tools may now, therefore, be made available to the general astronomical community on AstroGrid empowering users to cross associate the numerous datasets available and thereby extract new science.

However, due to the current developmental stage of the AstroGrid project using the association tools is not entirely straight forward. The applications require several input parameters regarding the data which are not currently readily available through AstroGrid. We suggest that the IVOA includes survey completeness limits and areas in their resource metadata standard which is currently being drawn up. It would also be helpful to include a recommended counterpart search radius based on the positional errors of the dataset contents. These metadata are also not queryable parameters at present and a user must discover the information via other means, we would advocate making such information directly accessible through the VO system.

A standard structure for VO association tools is also recommended which follows the pattern of the three applications developed here. Since all cross matching algorithms require similar parameters such as the source list, candidate list and magnitude histogram of the background catalogue, we suggest that these are routinely structured as VOTables when passed to a cross matching tool and the results returned in the same format. The user should be able to specify the number of matches per source to be returned and these to be sorted such that the most likely match appears first and each returned with an associated match probability.

At present some rather complex scripting is necessary to manipulate tables returned by the AstroGrid linked databases into the format required by the association tools. These tables must also be acquired via an http get request rather than a formal ADQL database query due to a limited number of ADQL commands being supported at present. However, AstroGrid is developing rapidly and in the near future these problems will be solved as more ADQL commands are enabled and the library used to manipulate tables in the scripts developed further.

We also implemented the Astronomical Distributed Annotation System (AstroDAS, Prina–Ricotti 2006) which stores the results of cross–association algorithms as mapping annotations in a database. Therefore, before embarking on the lengthy development and runtime of a cross–associating workflow users can search those that have already been done. The system is based on a distributed peer–to–peer architecture with each node of the network known as an 'annotation peer'. This allows users to publish and manage their own sets of cross–matched objects. We set up our own annotation peer and published to it the results of the HIPASS/SSA cross–matching using an application, `AnnotationsPopulator`, developed to convert the results of the three xmatch applications into the form required by AstroDAS and populate the annotation peer database. This application has been deployed on AstroGrid and is published in the registry allowing users to store their own cross–match results in their own annotation peer, it can be incorporated as the final step in a cross–matching workflow.

6.6 Conclusions and Future Work

As mentioned above the three association methods - the Mann et al. (1997, 2002) likelihood ratio, Poisson method and EM algorithm - may now be made available as tools on AstroGrid. Some documentation must be provided with these to describe the different methods and recommend under which circumstances to use each method. We suggest that the Poisson model, since it is a simple method, be used when choosing between only a handful of candidates per source. This method is computationally inexpensive and runs quickly, it could therefore be used as a first test and another method employed later. The EM algorithm was found to be a very effective association method, however, it is sensitive to the shape of the background magnitude histogram. We, therefore, recommend that the Mann et al. (1997, 2002) likelihood ratio method be used in cases where the background survey area is limited therefore producing a magnitude histogram with a restricted shape. Where associations are being sought for sources over a large area and the nature of the IDs required unknown the EM algorithm is preferable.

In the future many more association methods may be made available on AstroGrid, giving flexibility to users. In particular there is much to be learned from the informatics community regarding different machine learning techniques that may be very effective. The Rutledge et al. (2000) likelihood ratio technique could also be offered to be used in cases where IDs for over 1000 sources are being sought. This method also has the

advantage that it quantifies the cut–off probability for good associations which must be chosen somewhat arbitrarily in the Mann et al. (1997, 2002) likelihood ratio method. In the meantime AstroGrid and the VO are progressing in leaps and bounds and installing and using such tools will become ever simpler, empowering the entire astronomical community to gain as much knowledge as possible from the data available.

Bibliography

Abazajian, K. et al.: 2005, AJ, 129, 1755

Abel, D. J., Devereux, D., Power, R. A., Lamb, P. R.: 2004, CSIRO ICT Centre Technical Report, TR–04/1846

Babbedge, T. S. R. et al.: 2004, MNRAS 353, 654

Becker, R. H., White, R. L., Helfand, D. J.: 1995, ApJ 450, 559

Berriman, B., Kirkpatrick, D., Hanisch, R., Szalay, A., Williams, R.: 2003, IAUJD 8E, 60B

Bertin, E., Arnouts, S.: 1996, A&AS 117, 393

Bertin, E., Dennefeld, M., Moshir, M.: 1997, A&A 323, 685

Blain, A. W., Smail, I., Ivison, R. J., Kneib, J.-P.: 1999, MNRAS 302, 632

Bruzual A., G. and Charlot, S.: 1993, ApJ 405, 538

Burigana, C., Danese, L., de Zotti, G., Franceschini, A., Mazzei, P., Toffolatti, L.: 1997, MNRAS 287, L17

Cesarsky, C. J. et al.: 1996, A&A 315, 32C

Chapman, S. C., Smail, I., Ivison, R. J., Helou, G., Dale, D. A., and Lagache, G.: 2002, ApJ 573, 66

Chary, R. and Elbaz, D.: 2001, ApJ 556, 562

Ciliegi, P. et al.:1999, MNRAS 302, 222

Colless, M. et al.: 2001, MNRAS 328, 1039

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., Broderick, J. J.: 1998, AJ 115, 1693

Cram, L., Hopkins, A., Mobasher, B., and Rowan-Robinson, M.: 1998, ApJ 507, 155

Dempster, A. P., Laird, N. M., Rubin, D. B.: 1977, Journal of the Royal Statistical Society, Series B, 39:1-38

Dennefeld, M., Lagache, G., Mei, S., Ciliegi, P., Dole, H., Mann, R. G., Taylor, E. L., Vaccari, M.: 2005, A&A 440, 5

Dole, H. et al.: 1999, usis.conf.1031D

Dole, H. et al.: 2001, A&A 372, 364

Downes, A. J. B., Peacock, J. A., Savage, A., Carrie, D. R.: 1986, MNRAS 218, 31

Doyle, M. T. et al.: 2005, MNRAS 361, 34

Dunne, L., Eales, S., Edmunds, M., Ivison, R., Alexander, P., and Clements, D. L.: 2000, MNRAS 315, 115

Dunne, L. and Eales, S. A.: 2001, MNRAS 327, 697

Efstathiou, A. and Rowan-Robinson, M.: 2003, MNRAS 343, 322

Efstathiou, A., Rowan-Robinson, M., and Siebenmorgen, R.: 2000, MNRAS 313, 734

Elbaz, D., Cesarsky, C. J., Chanial, P., Aussel, H., Franceschini, A., Fadda, D., and Chary, R. R.: 2002, A&A 384, 848

Farrah, D., Afonso, J., Efstathiou, A., Rowan-Robinson, M., Fox, M., and Clements, D.: 2003, MNRAS 343, 585

Feigelson, E. D., Babu, G. J.: 2003, Statistical challenges in astronomy

Franceschini, A., Aussel, H., Cesarsky, C. J., Elbaz, D., and Fadda, D.: 2001, A&A 378, 1

González–Solares, E. et al.:2005, MNRAS 358, 333

Gray, J., Szalay, A.: 2002, Technical Report, MSR-TR-2002-75

Greve, T. R., Ivison, R. J., Bertoldi, F., Stevens, J. A., Dunlop, J. S., Lutz, D., Carilli, C. L.: 2004, MNRAS 354, 779

Hacking, P. B. and Soifer, B. T.: 1991, ApJ 367, 49

Hambly, N. C. et al.:2001, MNRAS 326, 1279

Hambly, N., Read, M., Mann, R., Sutorius, E., Bond, I., MacGillivray, H., Williams, P., Lawrence, A.: 2004, ASPC 314, 137

Hanisch, R. J., Farris, A., Greisen, E. W., Pence, W. D., Schlesinger, B. M., Teuben, P. J., Thompson, R. W., Warnock, A.: 2001, A&A 376, 359

Héraudeau, Ph. et al.: 2004, MNRAS 354, 924

Hildebrand, R. H., 2004, QJRAS, 24, 267

Ives, D. J., Tulloch, S., Churchill, J.: SPIE 2654, 266

Jarret, T. H., Chester, T., Cutri, R., Schneider, S. E., Huchra, J. P.: 2003, AJ 125, 525

Jarrett, T. H., Chester, T., Cutri, R., Schneider, S., Skrutskie, M., Huchra, J. P.: 2000, AJ 119, 2498

Kennicutt, R. C.: 1998, ARA&A 36, 189

Kessler, M. F. : 1987, MitAG 68, 186

Kessler, M. F. et al.: 1996, A&A 315, L27

Lagache, G., Dole, H., and Puget, J.-L.: 2003, MNRAS 338, 555

Lawrence, A. et al., 1999, MNRAS, 308, 897

Lemke, D. et al.: 1996, A&A 315, L64

Lonsdale, C. J., Hacking, P. B., Conrow, T. P., Rowan-Robinson, M.: 1990, ApJ 358, 60

Lucy, L. B.: 1974, AJ 79, 745

Madau, P., Pozzetti, L., and Dickinson, M.: 1998, ApJ 498, 106

Mann, R. G., et al. : 1997, MNRAS 289, 482

Mann, R. G., et al. : 2002, MNRAS 332, 549

Mathis, J. S., Mezger, P. G., and Panagia, N.: 1983, A&A 128, 212

McGlynn, T. A., et al.: 2004, ApJ 616, 1284

McMahon, R. G., Walton, N. A., Irwin, M. J., Lewis, J. R., Bunclark, P. S., Jones, D. H.: 2001, NewAR 45, 97M

Meyer, M. J., et al.: 2004, MNRAS 350, 1195

Monet, D. G., et al.: 2003, AJ 125, 984

Moshir, M., et al.: 1990, IRASF C, 0

Oliver, S. J. and Rowan-Robinson, M. and Saunders, W.: 1992, MNRAS 256, 15

Oliver, S. J.: 1998, IAUS 179, 112O

Padovani, P., AVO: 2004, AAS 205, 130.06

Partridge, R. B., Peebles, P. J. E.: 1967 ApJ 148, 377

Patris, J., Dennefeld, M., Lagache, G., Dole, H.: 2003, A&A 412, 349

Postman, M. et al.: 2005, ApJ 623, 721

D. Prina–Ricotti, Masters Thesis, Universit di Roma Tre, 2006

Puget, J.-L., Abergel, A., Bernard, J.-P., Boulanger, F., Burton, W. B., Desert, F.-X., and Hartmann, D.: 1996, A&A 308, L5+

Puget, J.-L. et al.: 1999, A&A 345, 29

Rohde, D. J., Drinkwater, M. J., Gallagher, M. R., Downs, T., Doyle, M. T.: 2005, MNRAS 360, 69

Rowan-Robinson, M. et al.: 2004, MNRAS 351, 1290

Rutledge, R. E. and Brunner, R. J. and Prince, T. A., Lonsdale, C.: 2000, ApJS 131, 335R

Sajina, A., Borys, C., Chapman, S., Dole, H., Halpern, M., Lagache, G., Puget, J.-L., and Scott, D.: 2003, MNRAS 343, 1365

Scott, D.: 2000, A&A 375L, 5S

Scott, S. E. et al.: 2002, MNRAS 331, 817

Siebenmorgen, R. and Kruegel, E.: 1992, A&A 259, 614

Silva, L., Granato, G. L., Bressan, A., Danese, L., 1998, ApJ, 509, 103

Soifer, B. T., Neugebauer, G., Houck, J. R.: 1987, ARA&A 25, 187

Stickel, M. et al.: 2000, A&A 359, 865

Storkey, A., Williams, C., Taylor, E. L., Mann, R. G.: 2005, Informatics Research Report 318

Sunyaev, R. A., Tinsley, B. M., Meier, D. L.: 1978, CompAp 7, 183

Sutherland, W., Saunders, W.: 1992, MNRAS 259, 413S

Taylor, E. L. et al.: 2005, MNRAS 361, 1352

Vaccari, M. et al.: 2005, MNRAS 358, 397

Wakamatsu, K., Colless, M., Jarrett, T., Parker, Q., Saunders, W., Watson, F.: 2003, ASPC 289, 97

Williams, R.: Grid Computing: Making The Global Infrastructure a Reality by Fran Berman, Anthony J.G. Hey, and Geoffrey Fox, Wiley, 2003, pp 837-858

Zwaan M. A.: 2004, MNRAS 350, 1210

A

FIRBACK N2 IDs

Notes on the multiwavelength associations for individual sources. Images of each of the sources are shown in Figure A.1

FN2 - 000: Identification with a bright optical galaxy (P_{ran} = 0.026, r magnitude = 18.5, no redshift). This is also an ELAIS source optical ID with detected emission at 15μm, 6.7μm, 90μm and radio detected emission at 20cm. There are two other ELAIS sources within the FIRBACK error ellipse with optical IDs with associated P_{ran} values of 0.604 and 0.246.

FN2 - 001: Identification with a bright optical galaxy (P_{ran} = 0.002, r magnitude = 18.3, z_{phot} = 0.10). This is also an ELAIS source optical ID with detected emission at 15μm, 6.7μm, 90μm and radio detected emission at 20cm. This also has an IRAS detection at 60μm.

FN2 - 002: Identification with an optical galaxy (P_{ran} = 0.078, r magnitude = 19.3). An ELAIS source ID has an optical ID with the same galaxy, however, they are different entries in the INT catalogue caused by the splitting up of bright objects by the image analyser. The ELAIS ID has P_{ran} = 0.086, z_{spect} = 0.17 and detected emission at 15μm, 6.7μm and 90μm. This also has an IRAS detection

at $60\mu m$. There are two other ELAIS sources within the FIRBACK error ellipse with optical IDs with associated P_{ran} values of 0.885 and 0.983.

FN2 - 003: Identification with a bright optical galaxy (P_{ran} = 0.001, r magnitude = 17.4, z_{spect} = 0.03). This is also an ELAIS source optical ID with detected emission at $15\mu m$, $6.7\mu m$, $90\mu m$ and radio detected emission at $20cm$. This also has an IRAS detection at $60\mu m$.

FN2 - 004: This ID is a small optical galaxy (P_{ran} = 0.079, r magnitude = 20.2). However, there is also present another large bright galaxy (r magnitude = 16.3) which is mistakenly classified as a star so is forcably not associated and assigned a P_{ran} of 1.0. Previous associations allowing FIRBACK sources to be associated with stars chose this ID which matches with the optical ID of an ELAIS source which has z_{spect} = 0.03 and detected emission at $15\mu m$, $6.7\mu m$, $90\mu m$ and radio emission at $20cm$. This also has an IRAS detection at $60\mu m$.

FN2 - 005: Identification with bright optical galaxy (P_{ran} = 0.011, r magnitude = 18.7, z_{spect} = 0.26). This is also an ELAIS source optical ID with detected emission at $15\mu m$, $6.7\mu m$, $90\mu m$ and radio emission at $20cm$. This also has an IRAS detection at $60\mu m$. There is another ELAIS source within the FIRBACK error ellipse with an optical ID with an associated P_{ran} of 0.315.

FN2 - 006: Identification with what appears on the image to be a pair of interacting galaxies (P_{ran} = 0.001, r magnitude = 18.2, no redshift). An ELAIS source ID has an optical ID with the same galaxy, however, they are different entries in the INT catalogue caused by the splitting up of bright objects by the image analyser. The ELAIS source has detected emission at $15\mu m$, $6.7\mu m$ and $90\mu m$ and radio emission at $20cm$. There is another ELAIS source within the FIRBACK error ellipse which does not have an optical ID.

FN2 - 007: Identification with bright optical galaxy (P_{ran} = 0.022, r magnitude = 17.9, z_{spect} = 0.12). This is also an ELAIS source optical ID with detected emission at $15\mu m$, $6.7\mu m$, $90\mu m$ and radio emission at $20cm$. This also has an IRAS detection at $60\mu m$. There is another ELAIS source within the FIRBACK error ellipse with an optical ID with an associated P_{ran} of 0.998.

FN2 - 008: Identification with an optical galaxy (P_{ran} = 0.053, r magnitude = 18.4). However, the $175\mu m$ flux has been assigned in the ELAIS catalogue to a source

with a different optical ID. This ID has $P_{ran} = 0.084$ for association with the FIRBACK source and is therefore a possible identification. The ELAIS source has $z_{spect} = 0.14$ and detected emission at $15\mu m$, $6.7\mu m$, $90\mu m$ and radio emission at $20cm$. This also has an IRAS detection at $60\mu m$.

FN2 - 009: This source does not have a confident optical identification. It is best associated with a galaxy with $P_{ran} = 0.440$, this does not agree with the optical ID for the ELAIS source ($P_{ran} = 0.880$, r magnitude=18.8) to which the $175\mu m$ flux has been assigned, this source also has detected emission at $15\mu m$, $6.7\mu m$, $90\mu m$ and radio emission at $20cm$. This source is confidently identified as a star when stellar associations are allowed with $P_{ran} = 0.142$ (the threshold for these associations is $P_{ran} = 0.4$), however, this still does not agree with the ELAIS optical ID.

FN2 - 010: Identification with, what appears to be from the image, a pair of interacting galaxies ($P_{ran} = 0.114$, r magnitude = 18.8). An ELAIS source ID has an optical ID with the same galaxy, however, they are different entries in the INT catalogue caused by the splitting up of bright objects by the image analyser. The ELAIS source has $z_{spect} = 0.17$ and detected emission at $90\mu m$ and radio emission at $20cm$. This also has an IRAS detection at $60\mu m$.

FN2 - 011: Identification with an optical galaxy ($P_{ran} = 0.118$, r magnitude = 19.8, $z_{phot} = 0.51$). However, the $175\mu m$ flux has been assigned in the ELAIS catalogue to a source with a different galaxy ID (r magnitude = 18.6) and detected emission only in the far-infrared and radio. This has a P_{ran} value of 0.547 for association with the FIRBACK source, it is therefore not an acceptable association.

FN2 - 012: Identification with an optical galaxy ($P_{ran} = 0.006$, r magnitude = 18.5). However, the $175\mu m$ flux has been assigned in the ELAIS catalogue to a source with a different galaxy ID (r magnitude = 20.0). This ID has $P_{ran} = 0.088$ for association with the FIRBACK source and is therefore a possible identification. The ELAIS source has $z_{spect} = 0.14$ and detected emission at $15\mu m$, $6.7\mu m$, $90\mu m$ and radio emission at $20cm$. There are two other ELAIS sources within the FIRBACK error ellipse with optical IDs with associated P_{ran} values of 0.990 and 1.000.

FN2 - 013: This source does not have a confident optical identification. It is best associated with an optical galaxy ($P_{ran} = 0.260$, r magnitude = 20.4). This is

also an ELAIS source optical ID with detected emission at $15\mu m$.

FN2 - 014: This source appears in the ELAIS $90\mu m$ and $175\mu m$ unassociated catalogue. It does not have a confident optical identification.

FN2 - 015: Identification with bright optical galaxy ($P_{ran} = 0.015$, r magnitude = 18.6, $z_{spect} = 0.17$). This is also an ELAIS source optical ID with detected emission at $15\mu m$, $6.7\mu m$, $90\mu m$ and radio emission at $20cm$.

FN2 - 016: Identification with an optical galaxy ($P_{ran} = 0.119$, r magnitude = 20.0). An ELAIS source ID has an optical ID with the same galaxy, however, they are different entries in the INT catalogue caused by the splitting up of bright objects by the image analyser. This source has $z_{spect} = 0.13$ and detected emission at $15\mu m$, $6.7\mu m$ and radio emission at $20cm$. This also has IRAS detections at $60\mu m$ and $100\mu m$. There are two other ELAIS sources within the FIRBACK error ellipse one with an optical ID with $P_{ran} = 1.000$, the other has no optical ID.

FN2 - 017: Identification with an optical galaxy ($P_{ran} = 0.030$, r magnitude = 18.5, $z_{phot} = 0.18$). The $175\mu m$ flux is assigned to an ELAIS source with a galaxy ID with associated $P_{ran} = 1.000$, it is very faint and far from the source. There is, however, another ELAIS source present, with detected emission only in the radio, in the error ellipse without an optical ID which is only about 5arcsec from our FIRBACK optical ID.

FN2 - 018: Identification with an optical galaxy ($P_{ran} = 0.037$, r magnitude = 19.5, $z_{phot} = 0.15$). However, the $175\mu m$ flux has been assigned in the ELAIS catalogue to a source with a different galaxy ID with $P_{ran} = 0.998$ (r magnitude = 20.5), it is therefore not a possible ID. This also has an IRAS detection at $60\mu m$. There is another ELAIS source within the FIRBACK error ellipse whose galaxy ID has a $P_{ran} = 1.000$

FN2 - 019: Identification with an optical galaxy ($P_{ran} = 0.001$, r magnitude = 18.0, $z_{phot} = 0.10$). The ELAIS source to which the $175\mu m$ flux has been assigned does not have an optical ID, it has detected emission at $6.7\mu m$ and $20cm$. This also has an IRAS detection at $60\mu m$.

FN2 - 020: Identification with an optical galaxy ($P_{ran} = 0.086$, r magnitude = 20.2, $z_{phot} = 0.15$). The ELAIS source to which the $175\mu m$ flux has been assigned

is fainter (r magnitude = 22.1) and further from the source than our ID and therefore has a higher P_{ran} value of 0.990.

FN2 - 021: Identification with an optical galaxy ($P_{ran} = 0.001$, r magnitude = 17.57, $z_{phot} = 0.10$). However, the flux of this FIRBACK source is below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

FN2 - 022: Identification with an optical galaxy ($P_{ran} = 0.019$, r magnitude = 19.1). The ELAIS source to which the $175\mu m$ flux has been assigned has an optical identification with a bright optical galaxy with r magnitude = 18.8, it has $P_{ran} = 0.140$ and is therefore a possible identification for the FIRBACK source. This ELAIS source has $z_{spect} = 0.17$ and detected emission at $15\mu m$, $6.7\mu m$ and radio emission at $20cm$. There is another ELAIS source within the FIRBACK error elipse, however, its optical ID is stellar and therefore it has $P_{ran} = 1.0$.

FN2 - 023: Identification with an optical galaxy ($P_{ran} = 0.018$, r magnitude = 19.04, $z_{phot} = 0.02$). However, the flux of this FIRBACK source is below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

FN2 - 024: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

FN2 - 025: Identification with an optical galaxy ($P_{ran} = 0.119$, r magnitude = 20.2, $z_{phot} = 0.20$).The ELAIS source to which the $175\mu m$ flux has been assigned has optical ID classed as a galaxy, however, it looks stellar and has a P_{ran} of 0.974 so is not a possible ID. There is another ELAIS source within the FIRBACK error elipse, however, its optical ID is stellar and therefore it has $P_{ran} = 1.0$. This also has IRAS detections at $60\mu m$ and $100\mu m$.

FN2 - 026: This source does not have a confident optical ID. It is associated with a galaxy with r magnitude 22.4 with a $P_{ran} = 0.906$. There is an ELAIS source within the FIRBACK error ellipse, however, it has a stellar optical ID and therefore has $P_{ran} = 1.0$.

CFN2 - 027: This source does not have a confident optical ID. It is associated with a galaxy with r magnitude 20.7 with a $P_{ran} = 0.305$. When allowing associations with all objects this source confidently associated with a star ($P_{ran} = 0.073$), however, this still does not agree with the ELAIS optical ID which is a galaxy with $P_{ran} = 0.999$.

CFN2 - 028: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 029: Identification with an optical galaxy ($P_{ran} = 0.070$, r magnitude = 19.6, $z_{phot} = 0.15$). The $175\mu m$ flux has been assigned to a fainter galaxy (r magnitude = 22.2) which is further from the FIRBACK position than our ID, it therefore has a high P_{ran} of 1.0. This also has an IRAS detection at $60\mu m$. There are three other ELAIS sources within the FIRBACK error ellipse, however, none of them have optical IDs.

CFN2 - 030: Identification with an optical galaxy ($P_{ran} = 0.086$, r magnitude = 20.26, $z_{phot} = 0.15$). However, the flux of this FIRBACK source is below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue. However, it has IRAS detections at $60\mu m$ and $100\mu m$.

CFN2 - 031: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 032: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 033: This source does not have a confident optical ID. It is associated with a galaxy with r magnitude 21.8 with a $P_{ran} = 0.934$.

CFN2 - 034: Identification with an optical galaxy ($P_{ran} = 0.030$, r magnitude = 19.42, $z_{phot} = 0.55$). However, the flux of this FIRBACK source is below the ELAIS

catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 035: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 036: Identification with an optical galaxy ($P_{ran} = 0.119$, r magnitude = 19.40, $z_{phot} = 0.10$). However, the flux of this FIRBACK source is below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue. However, it has IRAS detections at $60\mu m$ and $100\mu m$.

CFN2 - 037: This source does not have a confident optical ID. It is associated with a galaxy with r magnitude 22.1 with a $P_{ran} = 0.812$. When associations are allowed with all objects this source is confidently identified as a star with $P_{ran} = 0.183$. However, this still does not agree with the stellar optical ID of the ELAIS source assigned the $175\mu m$ flux.

CFN2 - 038: Identification with an optical galaxy ($P_{ran} = 0.023$, r magnitude = 19.3, $z_{spect} = 0.14$). This is also the optical ID for an ELAIS source with detected emission at $15\mu m$ and has an IRAS detection at $100\mu m$.

CFN2 - 039: Identification with an optical galaxy ($P_{ran} = 0.001$, r magnitude = 17.1, $z_{spect} = 0.07$). This is also the optical ID for an ELAIS source with detected emission at $15\mu m$, $6.7\mu m$, $90\mu m$ and radio emission at $20cm$. This also has an IRAS detection at $60\mu m$.

CFN2 - 040: This source does not have a confident optical ID. It is best associated with a galaxy with r magnitude 20.5 and $P_{ran} = 0.192$. When associations are allowed with all optical objects this source is confidently identified as a star with $P_{ran} = 0.278$. The ELAIS source to which the $175\mu m$ flux has been assigned has a galaxy optical identification with $P_{ran} = 0.562$. There is another ELAIS source within the FIRBACK error ellipse with a galaxy optical ID with $P_{ran} = 0.393$.

CFN2 - 041: This source does not have a confident optical ID, however, its best association agrees with the optical identification of the ELAIS source to which the $175\mu m$ flux has been assigned. This galaxy has r magnitude = 19.2 and

$P_{ran} = 0.311$, the ELAIS source has detected emission at $15\mu m$ and $20cm$. When associated with all optical objects this source is confidently associated with a star with $P_{ran} = 0.210$. There are two other ELAIS sources within the FIRBACK error ellipse, one has a galaxy optical ID with $P_{ran} = 0.998$ and the other does not have an optical identification.

CFN2 - 042: This source does not have a confident galaxy optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue. When associations are allowed with all objects this source is confidently identified as a star with $P_{ran} = 0.001$.

CFN2 - 043: This source does not have a confident optical identification when associations are only allowed with galaxies ($P_{ran} = 0.437$). However, when all objects are possible associations it is matched with a star ($P_{ran} = 0.056$) which is also the ELAIS source ID to which the $175\mu m$ flux has been assigned. This ELAIS source has detected emission at $6.7\mu m$.

CFN2 - 044: Identification with an optical galaxy ($P_{ran} = 0.053$, r magnitude = 20.0, $z_{phot} = 0.05$). The ELAIS source to which the $175\mu m$ flux has been assigned has a galaxy ID with a P_{ran} value of 0.995. This also has an IRAS detection at $60\mu m$. There is another ELAIS source within the FIRBACK error ellipse, however, this source does not have an optical ID.

CFN2 - 045: This source does not have a confident galaxy optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue. When associations are allowed with all objects this source is confidently identified as a star with $P_{ran} = 0.035$.

CFN2 - 046: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of $223mJy$ and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 047: Identification with an optical galaxy ($P_{ran} = 0.070$, r magnitude = 19.0). The $175\mu m$ flux has been assigned to an ELAIS source with a galaxy optical ID with $P_{ran} = 0.125$ which is therefore a possible association. This

source has $z_{spect} = 0.25$ and detections at $15\mu m$, $6.7\mu m$ and $20cm$. There is another ELAIS source within the FIRBACK error ellipse with a galaxy optical ID with $P_{ran} = 0.911$.

CFN2 - 048: This source does not have a confident optical identification, is is best associated with a galaxy with $P_{ran} = 0.920$. The ELAIS source to which the $175\mu m$ flux has been assigned does not have an optical identification.

CFN2 - 049: Identification with an optical galaxy ($P_{ran} = 0.092$, r magnitude = 20.3, $z_{phot} = 0.51$). An ELAIS source ID has an optical ID with the same galaxy, however, they are different entries in the INT catalogue caused by the splitting up of bright objects by the image analyser. The ELAIS source has detected emission at $15\mu m$ and $90\mu m$. This also has an IRAS detection at $60\mu m$. There is another ELAIS source within the FIRBACK error ellipse with a galaxy optical ID with $P_{ran} = 0.911$.

CFN2 - 050: This source falls between two WFS chips.

CFN2 - 051: This source does not have a confident optical identification, it is best associated with a galaxy with $P_{ran} = 0.866$. The ELAIS source to which the $175\mu m$ flux has been assigned has a stellar optical ID and therefore is not a possible ID for the FIRBACK source.

CFN2 - 052: This source does not have a confident optical identification, it is best associated with a galaxy with $P_{ran} = 0.466$. The ELAIS source to which the $175\mu m$ flux has been assigned has a galaxy optical ID with $P_{ran} = 0.962$. There are two other ELAIS sources within the FIRBACK error ellipse one with a stellar ID and one with a galaxy ID, however, both have $P_{ran} = 1.0$.

CFN2 - 053: This source does not have a confident optical identification, it is best associated with a galaxy with $P_{ran} = 0.441$. The ELAIS source to which the $175\mu m$ flux has been assigned has a galaxy optical ID with $P_{ran} = 0.989$.

CFN2 - 054: This source does not have a confident optical identification. When associated with all optical objects it is confidently associated with a star with $P_{ran} = 0.215$, the ELAIS source assigned the $175\mu m$ flux also has a stellar ID with $P_{ran} = 0.350$ which is below the limit of 0.4 and therefore a possible association.

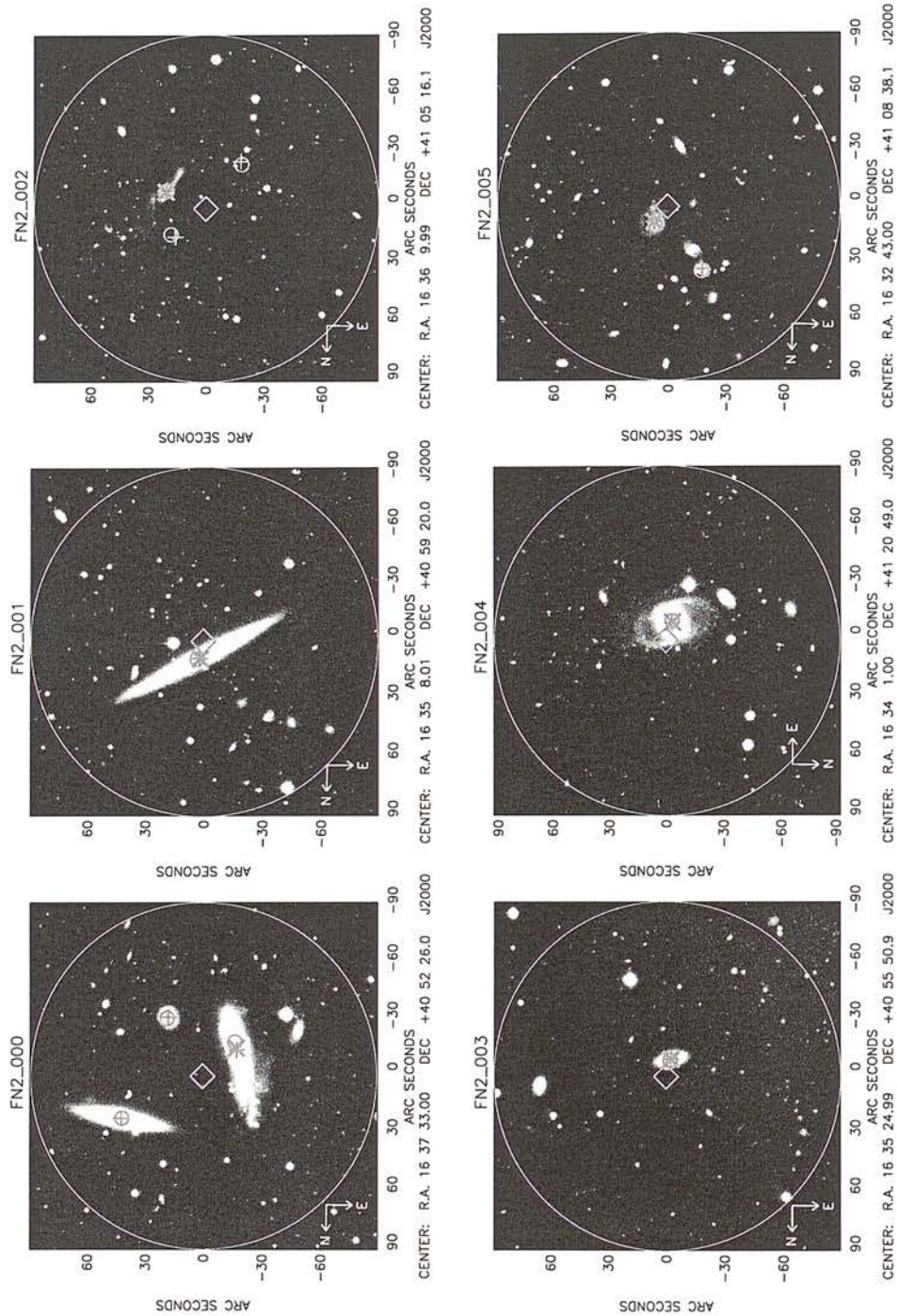

Figure A.1: Optical INT images of the N2 FIRBACK sources. FIRBACK source position marked with a diamond, optical ID marked with a cross, ELAIS positions marked with small circles, ELAIS optical IDs marked with pluses. A 90" circle is plotted to show the region in which optical IDs were searched for.

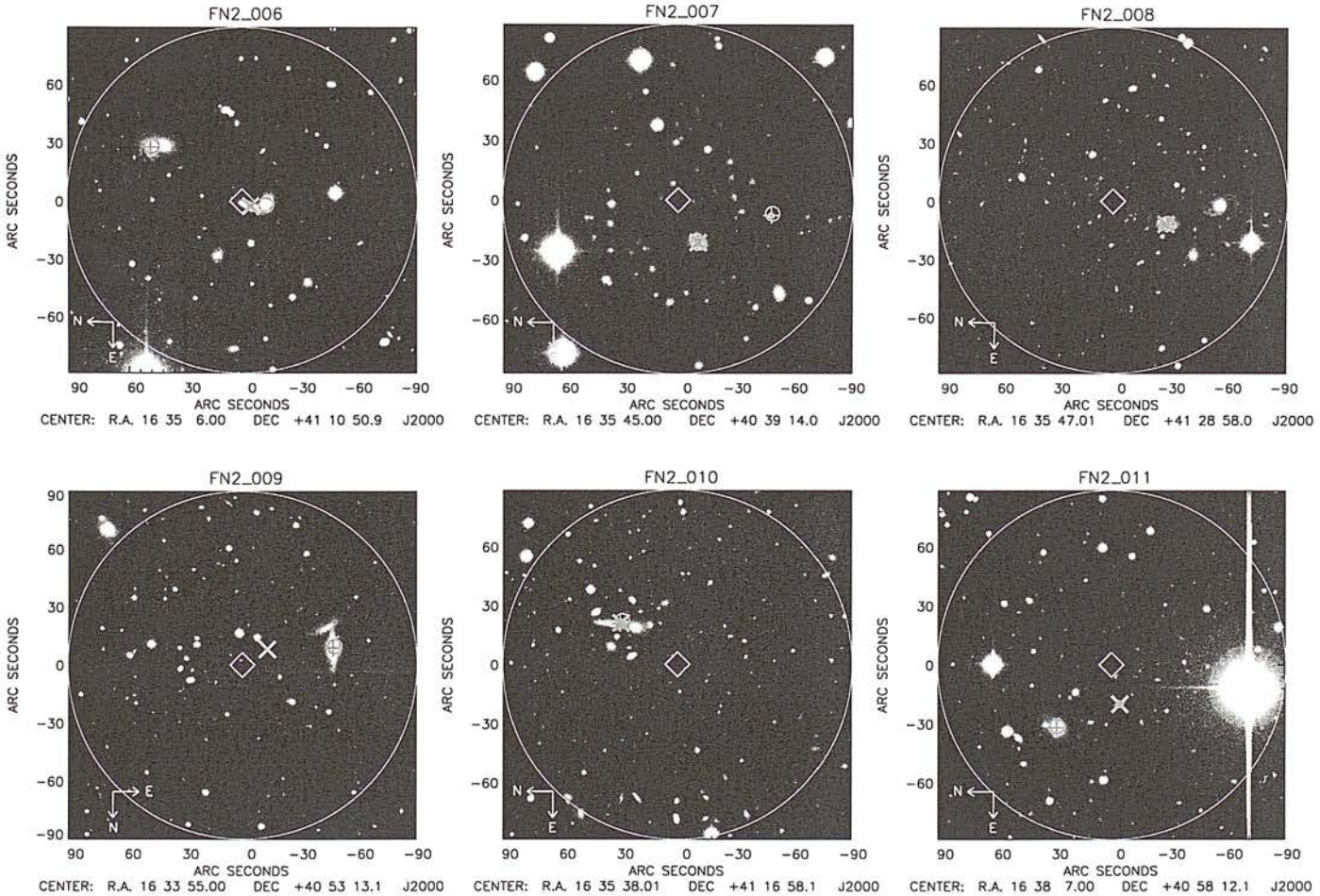

Figure A.1: *continued*

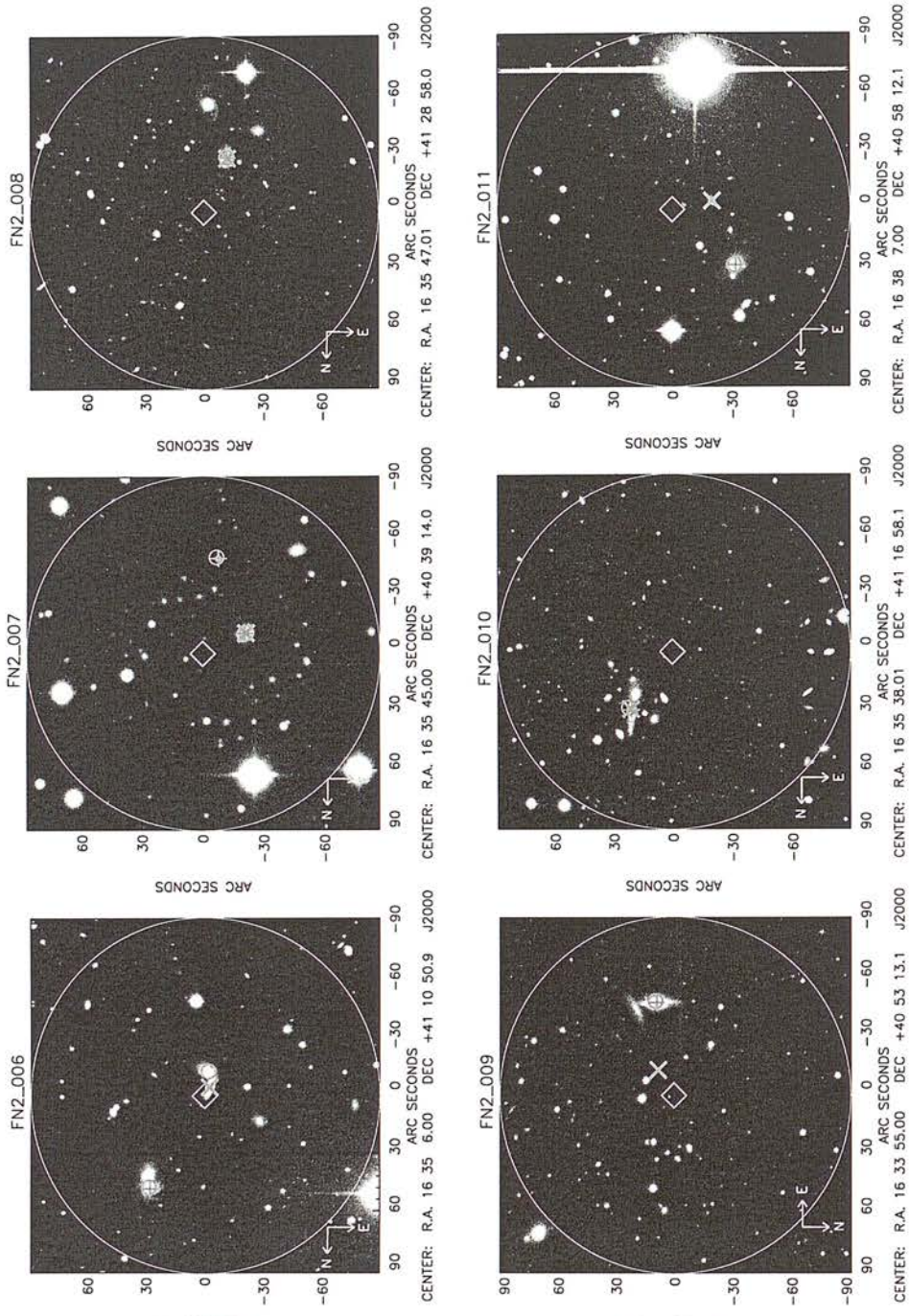

Figure A.1: *continued*

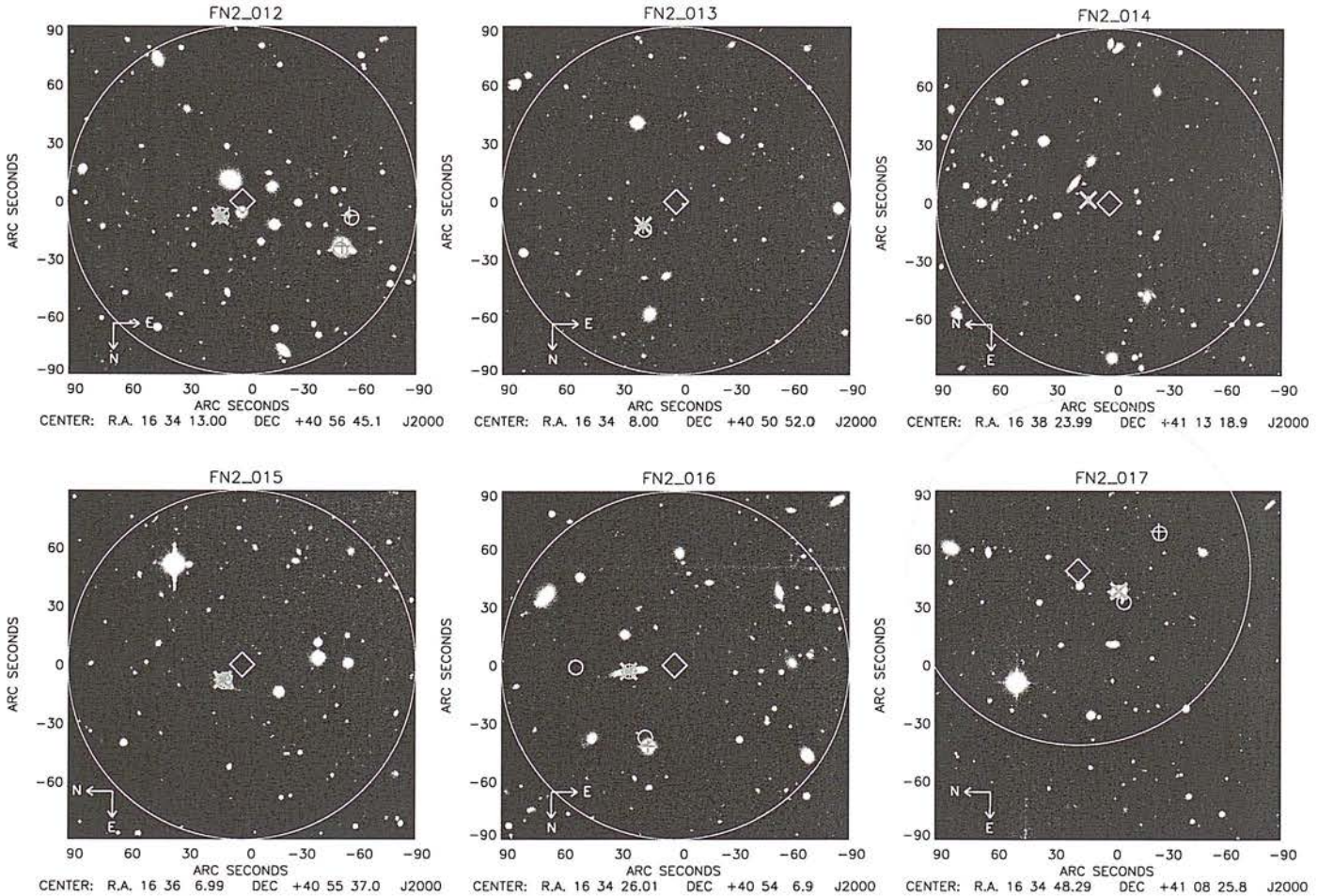

Figure A.1: *continued*

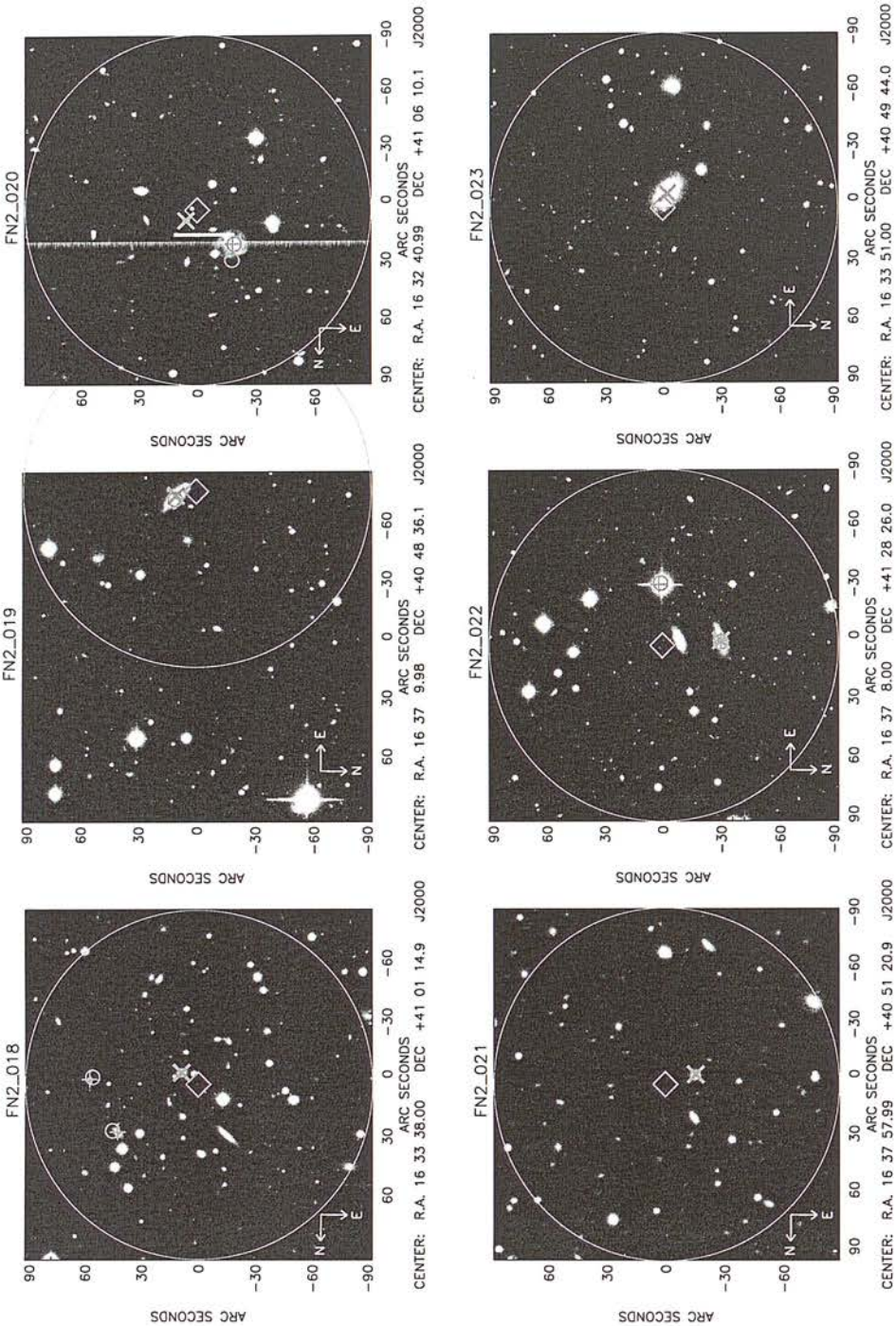

Figure A.1: *continued*

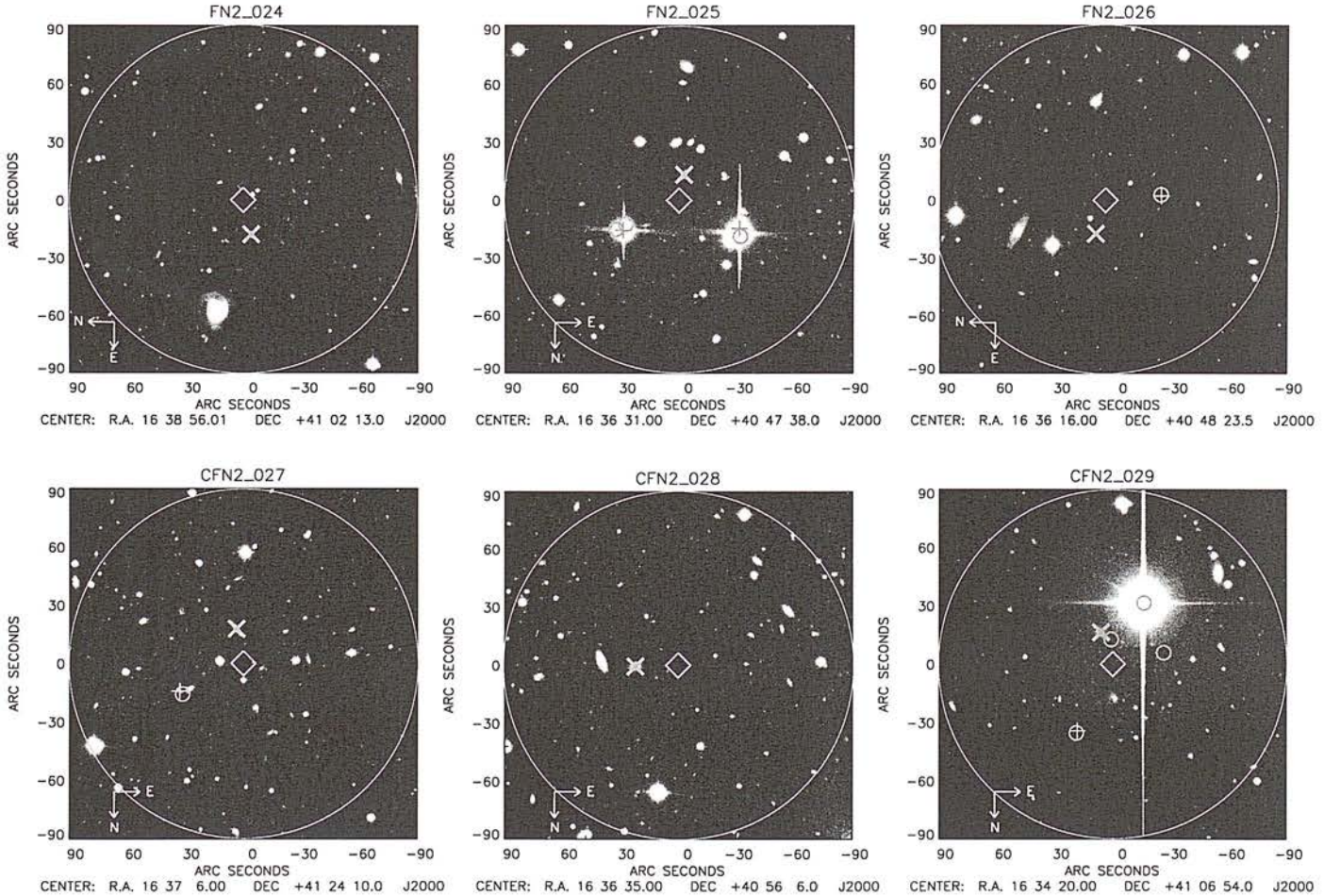

Figure A.1: *continued*

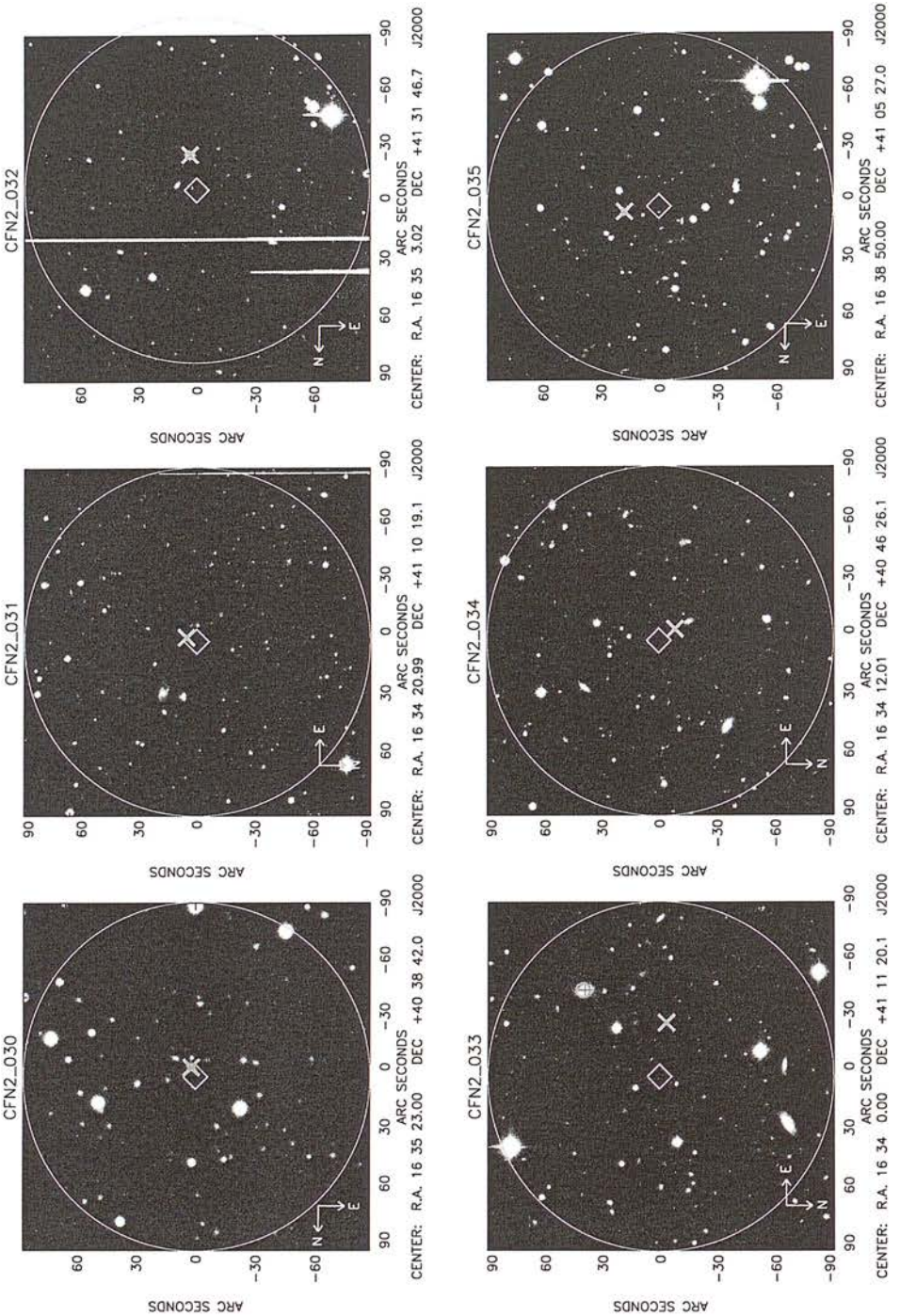

Figure A.1: *continued*

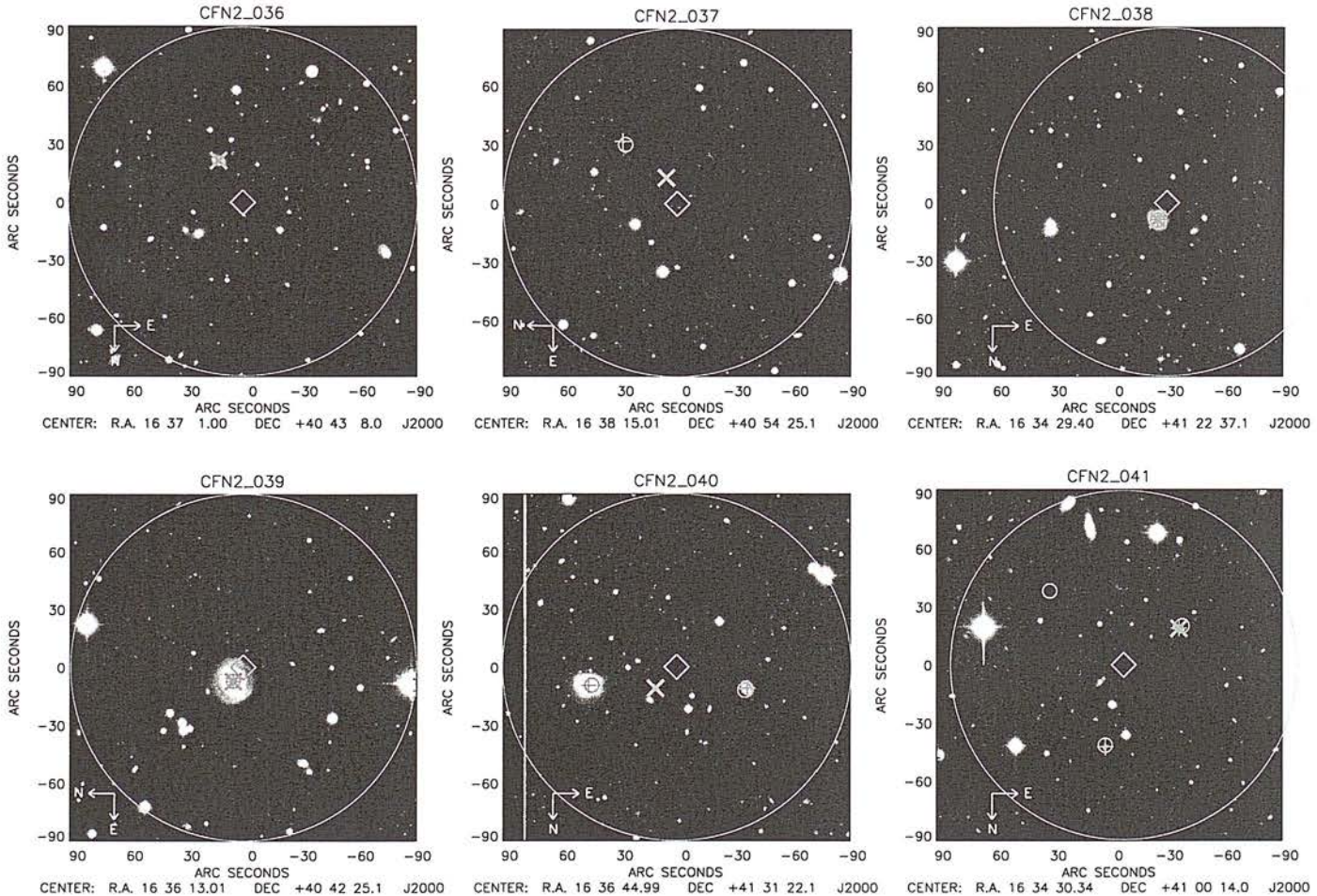

Figure A.1: *continued*

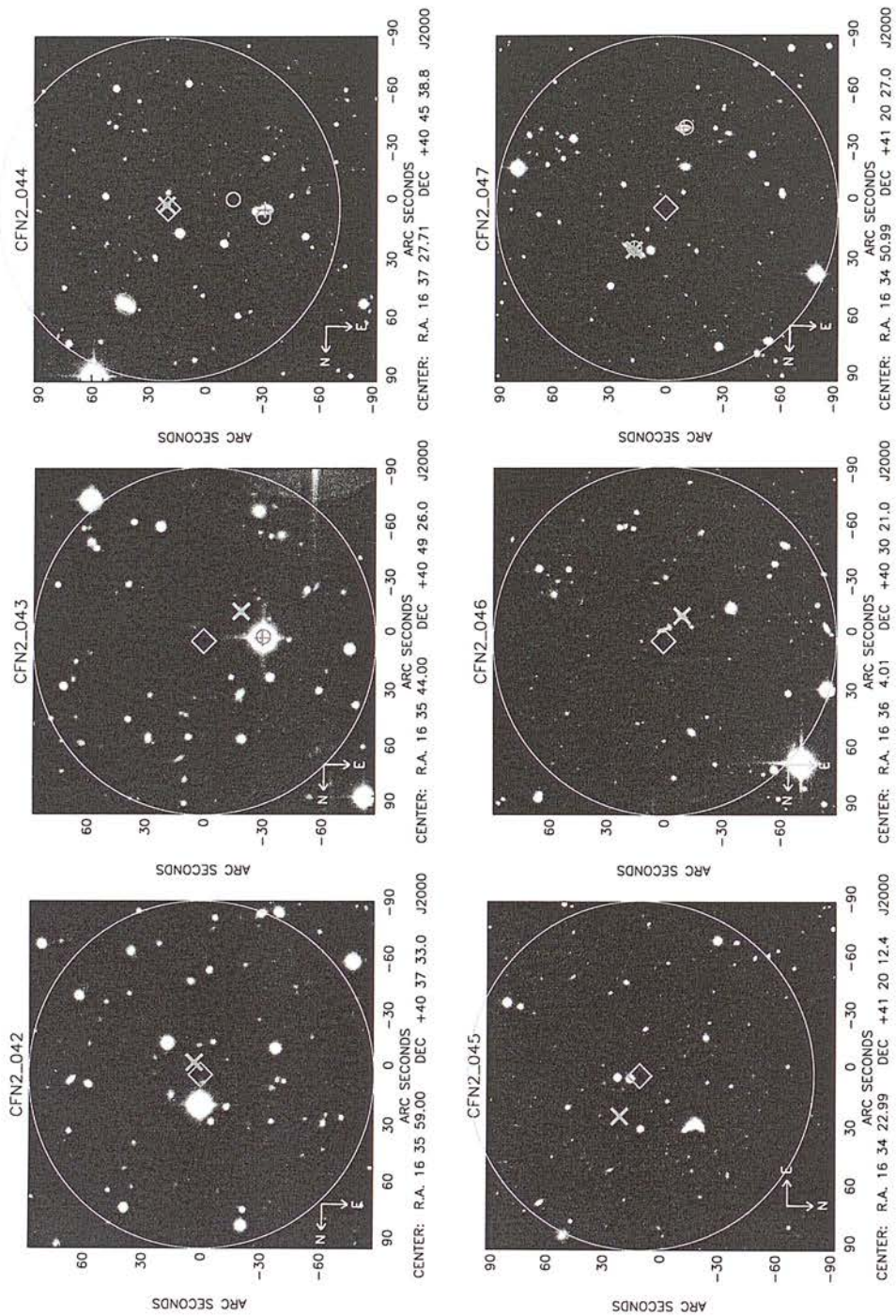

Figure A.1: *continued*

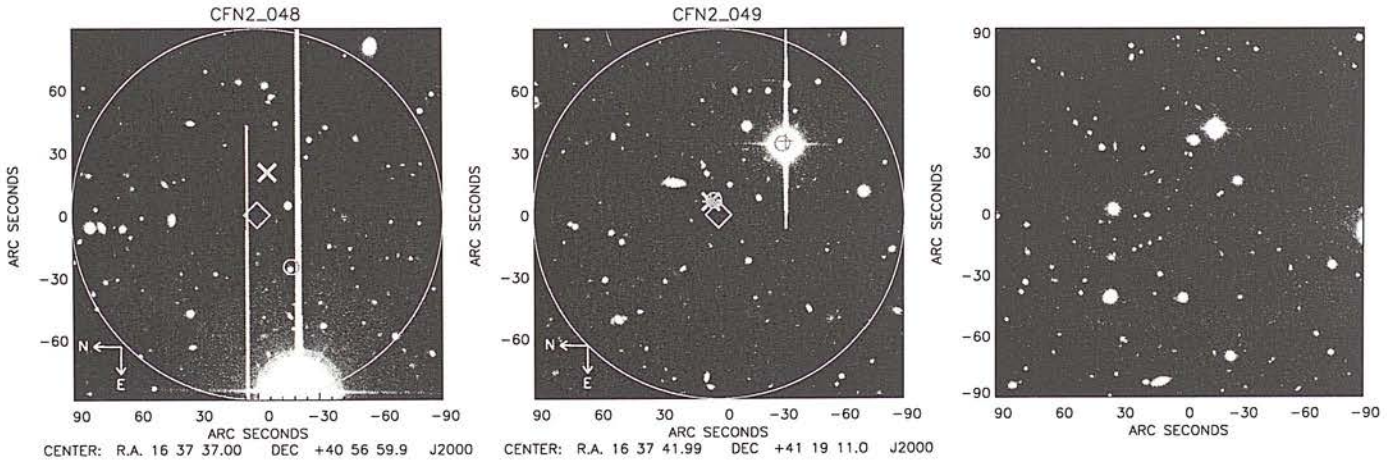

Figure A.1: *continued*

B

Copies of Published Papers

B.1 Published Papers

- Properties of FIRBACK–ELAIS 175-μm sources in the ELAIS N2 region (Taylor et al. 2005)

- FIRBACK IV. Towards the nature of the 170 μm source population (Dennefeld et al. 2005)

Mon. Not. R. Astron. Soc. 000, 000–000 (0000) Printed 18 April 2007 (MN LATEX style file v2.2)

Properties of FIRBACK-ELAIS $175\mu m$ sources in the ELAIS N2 region

E. L. Taylor[1], R. G. Mann[1], A. N. Efstathiou[2], T. S. R. Babbedge[3], M. Rowan-Robinson[3], G. Lagache[4], A. Lawrence[1], S. Mei[5], M. Vaccari[3], Ph. Héraudeau[6], S. J. Oliver[7], M. Dennefeld[8], I. Perez-Fournon[9], S. Serjeant[10], E. González-Solares[11], J.–L. Puget[4], H. Dole[12], C. Lari[13]

[1]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK*
[2]*Department of Computer Science and Engineering, Cyprus College, 6 Diogenes Str, 1516 Nicosia, Cyprus*
[3]*Astrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2BZ*
[4]*Institut d'Astrophysique Spatiale, Bât. 121, Université Paris XI, F-91405 Orsay Cedex*
[5]*Johns Hopkins University, 3400 N. Charles Street, 21218, Baltimore, MD, USA*
[6]*Kapteyn Astronomical Institute, P.O. Box 800, 9700 AV Groningen, The Netherlands*
[7]*Astronomy Centre, Department of Physics & Astronomy, University of Sussex, Brighton, BN1 9QJ, UK*
[8]*Institut d'Astrophysique de Paris, 98bis Boulavard Arago, F-75014 Paris*
[9]*Instituto de Astrofisica de Canarias, C/Vía Láctea s/n. 38200 La Laguna, Tenerife, Spain*
[10]*Centre for Astrophysics & Planetary Science, School of Physical Sciences, University of Kent, Canterbury, Kent CT2 7NR*
[11]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA*
[12]*Steward Observatory, University of Arizona, 933 N Cherry Ave Tuscon AZ 85721, USA*
[13]*Instituto do Radioastronomia, Via P. Gobetti 101, 40129, Bologna, Italy*

Received ??

ABSTRACT

We report on a search for the optical counterparts of 175 μm – selected sources from the Far–Infrared Background (FIRBACK) survey in the European Large Area ISO Survey (ELAIS) N2 field. Applying a likelihood ratio technique to optical catalogues from the Isaac Newton Telescope – Wide Field Survey (INT–WFS), we found optical identifications for 33 out of 55 FIRBACK sources in this field. These were then re-assessed in the light of associations with the ELAIS final catalogue for the the N2 field, to yield a final set of 31 associations. We have investigated the nature of this population through a comparison of their observed spectral energy distributions (SEDs) with predictions from radiative transfer models which simulate the emission from both cirrus and starburst components. We find the far–infrared sources to be 80 per cent star bursting galaxies with their starburst component at a high optical depth. The resulting SEDs were used to estimate far–infrared luminosities, star formation rates, dust temperatures and dust masses. The N2 FIRBACK population is found to consist of four suspected ultra–luminous infrared galaxies (ULIRGs) with $L_{FIR} \sim 10^{12}$ $L_{\odot}$ and $SFR_{FIR} > 100$ $M_{\odot} yr^{-1}$, a number of luminous infrared galaxies (LIRGs) with moderate star formation rates and $L_{FIR} \sim 10^{11}$ $L_{\odot}$ and a population of low redshift quiescently star forming galaxies. We also discuss the implications of these results for current evolutionary models.

Key words: galaxies: evoluntion – galaxies: fundamental parameters – galaxies: starburst – infrared: galaxies – surveys

1 INTRODUCTION

Evidence available in the 1960s and 1970s from regions of active star formation in nearby galaxies, and our own Milky Way, suggested that star forming galaxies would be strong sources of thermal infrared radiation, and implied that regions of dust and molecular clouds should be found within them (Sunyaev et al. 1978). It was then predicted that this dust reprocessed redshifted starlight from distant galaxies, would form a far–infrared background (FIRB) signal (Partridge and Peebles 1967).

With the advent of the Infrared Astronomical Satellite (IRAS) in the 1980s hundreds of previously undetected

galaxies were discovered, which emit up to 95 per cent of their total luminosity in the infrared (Soifer et al. 1987). This indicated significant dust reprocessing of their starlight and implied that star formation rates previously calculated from rest frame optical/uv luminosities were only a lower limit. Follow up redshift surveys, such as QDOT (Lawrence et al. 1999), showed that these galaxies were far more numerous in the past (e.g. Oliver et al. 1992), requiring stronger cosmological evolution than had previously been predicted by conventional models for the passive evolution of galaxies and providing support was building for models in which starbursts are important evolutionary mechanisms (Lonsdale et al. 1990).

The IRAS results allowed far tighter limits to be put on the level of the predicted far–infrared background. Hacking and Soifer (1991) predicted that the background should be detectable by the Cosmic Background Explorer satellite (COBE) at 100 μm and noted that discovering its exact level would allow further refinement of evolutionary models. No-evolution models placed a lower limit on the predicted background level, while detection of a strong background would indicate the existence of a population of objects with significant cosmological evolution. The FIRB was finally discovered by Puget et al. (1996) in the whole sky survey taken by FIRAS on COBE just as Hacking and Soifer (1991) had predicted.

Puget et al. (1996) found a signal which could be explained by the thermal emission from particles with temperatures of 20 and 12 K with emissivities of 3×10^{-6} and 5 K with an emissivity of 3×10^{-5}. This detection has fueled a new generation of galaxy evolution models but it is clear that to further differentiate between them we must understand the nature of the objects which contribute to the far–infrared background.

1.1 The Far–Infrared Background

The FIRBACK survey was carried out jointly with the ELAIS team at 175 μm using ISOPHOT (Lemke et al. 1996) on the Infrared Space Observatory (ISO) (Kessler et al. 1996). This survey aimed to resolve the FIRB into discrete sources and also allow studies of FIRB fluctuations. FIRBACK covered 3 main fields [South Marano, N1 and N2] covering a total area of 4 deg^2. The final source catalogue contains 106 sources with fluxes between 180 mJy and 2.4 Jy and a supplementary catalogue containing 90 sources with fluxes 135-180 mJy (Dole et al. 1999, 2001). The northern fields lie within the previously defined ELAIS N1 and N2 survey regions whose selection was based primarily on their low Galactic cirrus emission (as quantified by IRAS 100 μm emission) and high visibility, given ISO's orbit. This choice of fields allows multiwavelength follow–up thus providing a global view of galaxy evolution.

There have been several suggestions as to the nature of these 175 μm ISO sources. Two sources were examined in detail in the optical and near-IR by Chapman et al. (2002) who found two nearby ($z < 1$) galaxies with characteristic temperatures of ~30 K and ~50 K and bolometric luminosities (40–200 μm) of $10^{11} - 10^{12}$ L$_\odot$. These galaxies were classified as ULIRGs with morphologies suggesting the early stages of mergers adding weight to arguments for interactions sparking periods of increased star formation. A sample of ISO

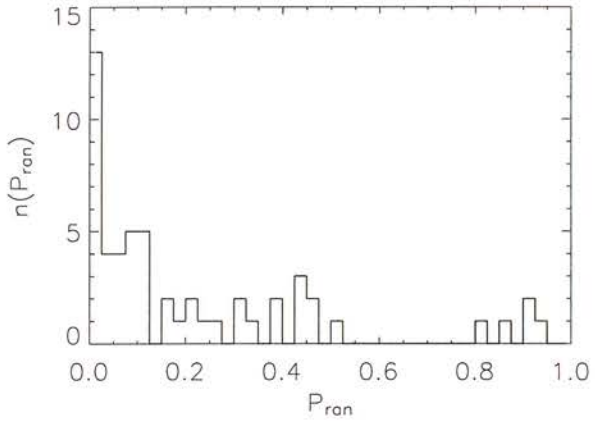

Figure 1. Distribution of P_{ran} values for chosen source counterparts.

175 μm sources were also examined at sub–mm and near-IR wavelengths by Sajina et al. (2003) who found a bimodal galaxy population: one population of normal star forming galaxies at $z \sim 0$ and a second more luminous population at $z \sim 0.4 - 0.9$. There has been spectroscopic follow–up of the brightest 175 μm sources in the FIRBACK South Marano field (Patris et al. 2003) which were found to be nearby ($z < 0.3$), heavily extincted ($A_v \sim 3$ with extreme cases up to $A_v = 7$), star forming galaxies with moderate star formation rates (a few 10 M$_\odot$yr^{-1}). These were classified at LIRGs, with $L_{IR} \simeq 10^{11}$ L$_\odot$.

We have gone one step further by seeking associations for all 55 N2 175μm ISO sources in the FIRBACK survey with optical sources from the INT–WFS catalogues in four bands (g',r',i',Z), the other ELAIS survey wavelengths (J, H, K, 15, 6.7, 90 μm, 20 cm) and IRAS wavelengths (60, 100 μm). (Also see Dennefeld et al. 2005 for investigations into sources in the N1 field). In doing so we are able to gain a multiwavelength view of the nature of these interesting objects. We have compared their SEDs with those predicted by the radiative transfer models of Efstathiou and Rowan-Robinson (2003) to provide further insight into their properties.

The layout of this paper is as follows: Section 2 describes the multiwavelength data, Section 3 discusses the methods of associating the multiwavelength data with the 175 μm ISO sources, provides a summary of the results of this process and briefly describes the optical properties of the sources. The radiative transfer models are described in Section 4 and Secton 5 describes the comparison of the sources' SEDs with predictions from the models. The resulting star formation rates, far–infrared luminosities and dust temperature estimations are given in Sections 6 and 7. Section 8 provides a summary of current evolutionary models, while conclusions and a discussion are provided in Section 9. An Appendix provides details of individual associations.

2 COMPLEMENTARY DATA

Optical associations were sought in r' data from the INT–WFS (McMahon et al. 2001) which used the Wide Field

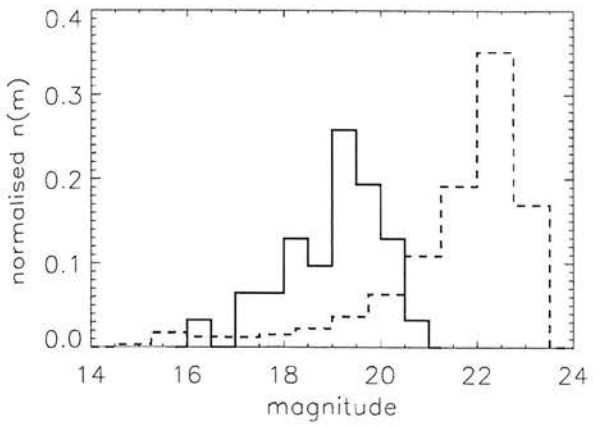

Figure 2. Normalised magnitude distribution of the N2 175 μm ISO source optical counterparts (r'-band) (solid line), dashed line gives the r'-band magnitude distribution of the INT-WFS catalogue for comparison.

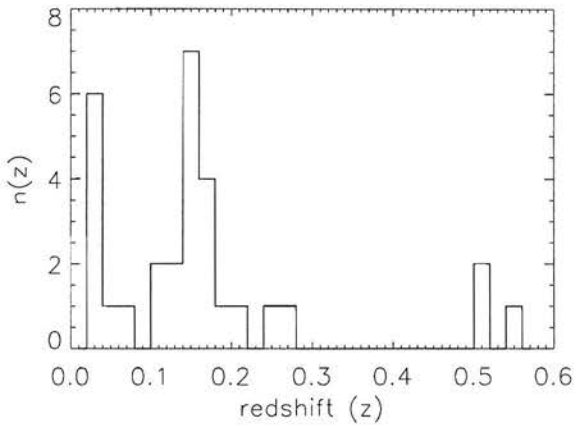

Figure 3. Redshift distribution for the 30 175 μm sources with spectroscopic or photometric redshifts as listed in Table 1.

Camera (WFC) on the 2.5 m Isaac Newton Telescope at the Observatorio del Roque de Los Muchachos (La Palma, Spain). The WFC is formed by four 4k x 2k CCDs. The arrays have 13.5 μm pixels corresponding to 0.33 arcsec/pixel at the telescope prime focus and each one covers an area on the sky of 22.8 arcmin x 11.4 arcmin. The total sky coverage per exposure for the array is therefore 0.29 deg^2. Gaps between detectors are typically 20 arcsec and optical observaions are carried out allowing for a 10 per cent overlap between adjacent pointings for photometric purposes. The ELAIS-N2 optical field is approximately 9 deg^2. The survey has a typical completeness limit of 23 mag (Vega) in wavebands U, g', r', i' and Z. However, there are no U-band data for our sources the N2 field. All optical magnitudes were calculated directly from the INT-WFS data.

Associations were then sought with sources in the multiwavelength ELAIS band-merged catalogue at 6.7, 15 and 90 μm with associated data at U,g',r',i',Z,J,H,K and 20 cm (Rowan-Robinson et al. 2004) which by then had become available. This was a survey carried out using ISO covering a total of 12 deg^2 over five main fields, three in the north (N1, N2, N3) and two in the south (S1,S2). These areas were also surveyed at 20 cm with the Very Large Array (VLA) and AT. The separate wavebands making up the ELAIS survey each comprise an independent survey, the band-merged catalogue simply brings together the results of each of their final anayses, as described in Section 3.

Where detections were not found at 15 and 90 μm upper limits were extracted directly from the survey maps (Vaccari et al. 2004; Héraudeau et al. 2004) to a 3σ level using aperture photometry with an aperture size of 6 arcsec and 90 arcsec respectively. An aperture correction was applied to the 15 μm upper limits since 40 per cent of the point spread function (PSF) lies outside the aperture.

IRAS 100 and 60 μm fluxes or upper limits (3σ) were obtained at the galaxy optical positions using the IRAS Scan Processing and Inegration (SCANPI) internet facility[1] to provide extra constraints on the source SEDs.

Associations were also sought in the Submillimetre Common User Bolometer Array (SCUBA) 850 μm (Scott et al. 2002) and Max-Planck Millimeter Bolometer array (MAMBO) 1200 μm (Greve et al. 2004) catalogues within 10 and 3 arcsec of the optical position of the 175 μm counterpart, respectively. However, none were found for any of our 175 μm sources.

3 IDENTIFICATIONS: METHOD AND RESULTS

The positional accuracy of the ISO 175 μm sources was determined using simulated maps by Dole et al. (2001). Since 93 per cent of artificial sources brighter than 180 mJy in the simulated maps were recovered within a radius of 50 arcsec, associations were sought for the N2 sources within a 90 arcsec radius. This ensured the capture within the source error circle of the true source position for even faint source IDs. Associations were made using a likelihood ratio technique described in full in Mann et al. (2002). In brief, the likelihood ratio is the ratio of the probability of finding the true counterpart to a particular source at the source position with that magnitude to the probability of finding an object of the same magnitude there by chance. In terms of flux this takes the form:

$$\mathrm{LR} = \frac{\mathrm{q(f)e(x,y)}}{\mathrm{n(f)}} \qquad (1)$$

where q(f) is the flux probability distribution function for the source counterparts, e(x,y) is the probability distribution of positional offsets between the source and object and n(f) is the surface density of objects per unit flux interval. e(x,y) was taken to be a gaussian distribution with $\sigma = 20$ arcsec. This choice prevents too many objects being assigned moderate likelihood ratios, without excluding objects which should be considered or including objects which should not. The function q(f) is unknown for a new source population, as is the case here, and, as shown in Mann et al. (1997), an empirical estimate is very noisy for a small number of sources, q(f) was taken to be a constant. The probability, P_{ran}, that

[1] http://irsa.ipac.caltech.edu/applications/IRAS/Scanpi/

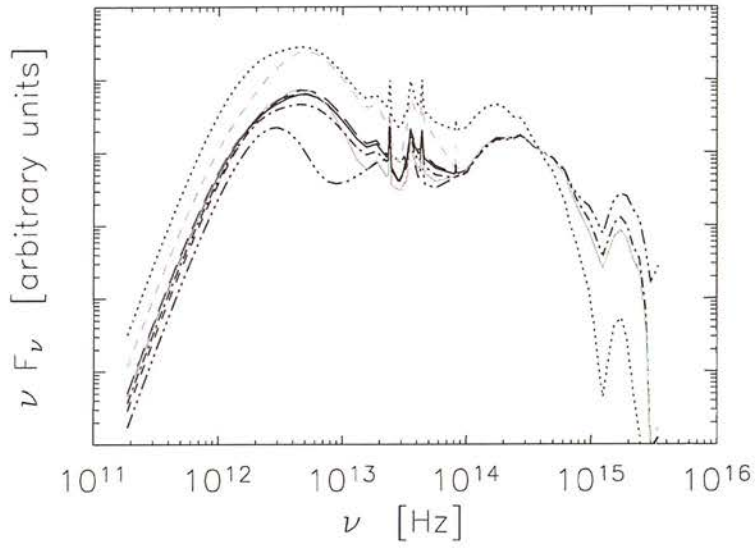

Figure 4. Examples of the affects of the radiative transfer model parameters on the resulting SEDs, see Section 4. Solid black line: $A_V = 1$, $f_{SB} = 1$, $\tau_v = 100$ (comparison model). Black dotted line: $A_V = 3$, $f_{SB} = 1$, $\tau_v = 100$. Dash-dot-dot-dot line: $t_m = 0$. Grey dashed line: $A_V = 1$, $f_{SB} = 5$, $\tau_v = 100$. Dashed line: $\psi = 10$. Dash-dot line: $\tau = 8 \times 10^{10}$. Long dashed line: $\tau_{SB} = 80$ Myrs. Solid grey line: $\tau_v = 200$.

a fictitious, randomly-placed source would have a likeliest assocation with an optical object in the catalogue producing a likelihood ratio at least as high as the source-object pair in question was then calculated. Thus good associations have low values of P_{ran}, corresponding to a high probability that the association is correct. The object with the lowest P_{ran} was chosen as the source optical ID, with a cut off for non-association of $P_{ran} = 0.15$. This is an arbitrary value chosen by inspection of the P_{ran} distribution, see Figure 1. It is expected (Mann et al. 2002) that true IDs will cluster at low P_{ran} values with a tail of objects reaching to high P_{ran} values due to sources without good associations in the optical catalogue, as seen in the plot.

Counterparts were also sought in the band-merged ELAIS multiwavelength catalogue (Rowan-Robinson et al. 2004) using a search radius of 60 arcsec consistent, with the band merging procedure used in the collation of this catalogue. The individual wavelength ELAIS catalogues were merged, taking into account the different positional accuracies of each catalogue, sequentially as follows. The 15 μm and 20 cm catalogues (Vaccari et al. 2004; Ciliegi et al. 1999) were first identified with optical objects from the INT–WFS catalogue (González–Solares et al. 2003) and then merged according to their optical positions using a search radius of 2 arcsec. These were then matched with the 6.7 μm catalogue using a search radius of 5 arcsec and those 6.7 μm sources not merged were associated with optical counterparts. These were then merged with the 90 μm and 175 μm sources (Héraudeau et al. 2004; Dole et al. 2001) using search radii of 30 arcsec and 60 arcsec respectively. Where these were matched with more than one catalogue entry the less probable assocations were flagged in the catalogue. After this those sources that were still unmatched were associated with optical counterparts. These optical counterparts were sought by Rowan-Robinson et al. (2004) independently from our work and the resulting optical IDs were then compared.

In summary, we sought associations on the basis of applying the likelihood ratio method to the optical data alone, then, with the availablility of the final band-merged ELAIS catalogue, we reassessed them. Our final optical and ELAIS wavelength ID list was drawn up using the following criteria:

(i) If our optical ID, selected using the likelihood ratio method, agreed with the ELAIS optical ID to which the 175 μm flux had been assigned then both the ELAIS optical ID and band-merged data were accepted.

(ii) If the ELAIS optical ID had a $P_{ran} < 0.15$ (cut off value see earlier) for association with the 175 μm ISO source then the ELAIS optical ID and band-merged data were accepted.

(iii) If our optical ID and ELAIS optical ID were in fact the same galaxy but had different entries in the INT–WFS catalogue due to the way the image analyser splits up bright galaxies into multiple sources, then the band-merged ELAIS data and optical ID were accepted.

(iv) If our optical ID had $P_{ran} < 0.15$ but no optical ELAIS catalogue match then the source has optical and 175 μm data only (it may also have IRAS detections/upper limits and 60 μm and 90 μm upper limits)

(v) If our optical ID had $P_{ran} > 0.15$ the source has no association

For sources in the main N2 FIRBACK catalogue ($S >$ 180 mJy), which contains 27 detections, we made 22 confident optical associations with galaxies, 7 of which agree exactly with the ELAIS optical ID (category i.) and 6 for which we accepted the ELAIS optical ID (categories ii. or iii.). Nine confident optical associations were made for sources in the supplementary catalogue ($S <$ 180 mJy, 28 sources), 2 of

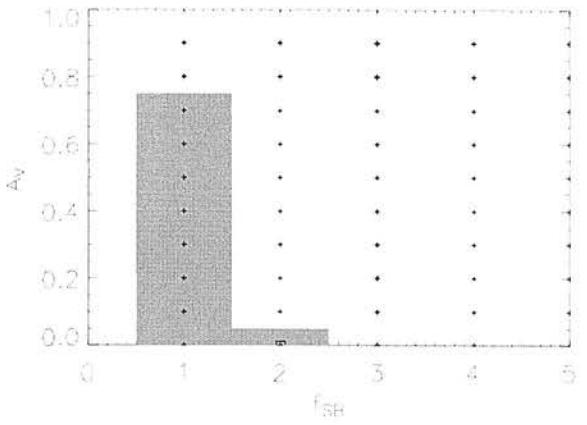

Figure 5. 1σ starburst model parameter confidence region for f_{SB} and A_V for FN2 008. Possible models parameter combinations are shown with small crosses (note $f_{SB} = 1 - 100$ and $A_V = 0 - 3$ were allowed), the best-fitting model is marked with a square and the 1σ confidence region is shaded. **A high resolution version of this figure is available at http://www.roe.ac.uk/~elt/N2paper_images.html .**

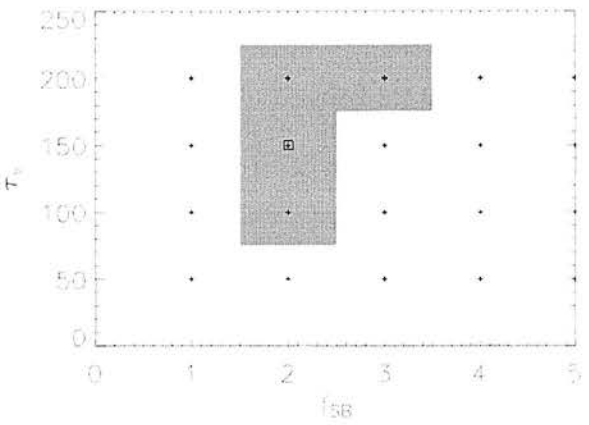

Figure 6. 1σ starburst model parameter confidence region for f_{SB} and τ_v for FN2 008. Possible models parameter combinations are shown with small crosses (note $f_{SB} = 1 - 100$ and $\tau_v = 50 - 200$ were allowed), the best-fitting model is marked with a square and the 1σ confidence region is shaded. **A high resolution version of this figure is available at http://www.roe.ac.uk/~elt/N2paper_images.html .**

which agreed with the ELAIS optical ID (category i.) and 2 of which where the ELAIS optical ID was accepted (categories ii. or iii.). There also were two sources in the supplementary catalogue with stellar IDs, one of which agreed with the ELAIS optical ID (category i.) and one of which was an accepted ELAIS optical ID (category ii. or iii.). The data for the identified sources are given in Table 1. and notes on individual sources are given in the Appendix, including P_{ran} values of any other plausible identifications.

The identified sources were morphologically classified by visual inspection. We used i'-band images from the INT-WFS, and followed the same classification method as Postman et al. (2005). The galaxies were classified using the classical Hubble sequence: E, E/S0, S0, S0/a, Sa, Sa/b, Sb, Sb/c, Sc, Irr. For the Advanced Camera for Surveys (ACS) on HST, for galaxies with redshift of around unity, visual classifications of this kind have a typical random error of 25 per cent, this reduces to 6 per cent when galaxies are classified in two broad cathegories: 1. early-type, 2. spirals and irregulars. These errors were estimated from the average scatter found in classifications by four ACS team members. The majority of our sources are spiral galaxies, we also have a number of S0 galaxies, 2 sets of interacting galaxies, 2 pairs and 2 irregulars, the morphology of 6 sources was unable to be deduced from the images.

Figure 2 shows the optical (r'-band) magnitude distribution of the N2 175 μm ISO source optical counterparts compared to that of the overall INT-WFS catalogue. The source IDs are, in general, bright objects with apparent magnitudes ranging from 21 to as bright as 16.

Figure 3 shows the redshift distribution of the sources: they lie predominantly at $z < 0.3$ with only two sources having $z > 0.5$. Redshifts are discussed in more detail in Section 5.

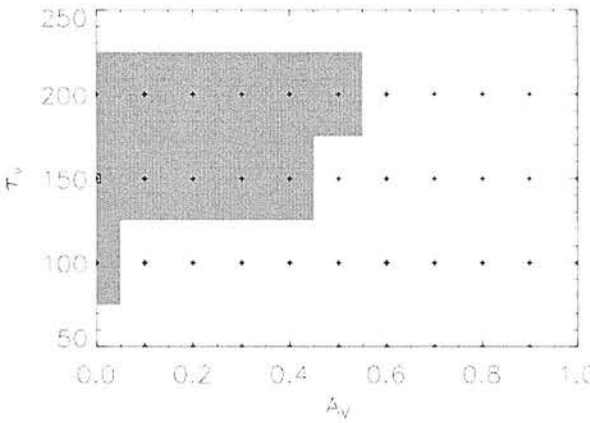

Figure 7. 1σ starburst model parameter confidence region for A_V and τ_v for FN2 008. Possible models parameter combinations are shown with small crosses (note $A_V = 0 - 3$ and $\tau_v = 50 - 200$ were allowed), the best-fitting model is marked with a square and the 1σ confidence region is shaded. **A high resolution version of this figure is available at http://www.roe.ac.uk/~elt/N2paper_images.html .**

4 RADIATIVE TRANSFER MODELS

It was initially attempted to fit the data to the GRASIL model SEDs of Silva et al. (1998). These provide six templates for nearby galaxies, three starbursts (Arp 220, M82 and NGC 6090) and three sprials (M100, M51 and NGC 6946). However, these templates did not describe the data well. Instead radiative transfer models of Efstathiou and Rowan-Robinson (2003) and Efstathiou et al. (2000) were employed to generate SED predictions for comparison with the data. These are able to

Table 1. Multiwavelength data for associated sources, fluxes in mJy.
nb. Sources listed with P_{ran} greater than the threshold 0.15 due to falling into category (iii) (see Section 3)
FN2 004 has P_{ran} = 1.0 since the associated galaxy is mistakenly classified as a star in the INT WFS catalogue.

FIRBACK	FIRBACK position	optical position	175 μm	P_{ran}	r'-mag	15 μm	6.7 μm	60 μm	90 μm	100 μm	20 cm	z_{phot}	z_{spec}	morphological classification
FN2 000	16 37 33 +40 52 26	16 37 34.53 +40 52 11.2	2377±213	0.026	18.46	53.9±0.3	18.3±0.1	-	1461±36	-	8.74±0.02	-	0.03	irregular or edge-on spiral with tidal tail
FN2 001	16 35 08 +40 59 20	16 35 07.87 +40 59 28.9	1251±139.	0.002	18.28	26.1±0.4	4.11±0.12	0.15±0.03	337±15	-	0.76±0.04	0.10	0.03	edge-on Sb/c
FN2 002	16 36 10 +41 05 16	16 36 08.15 +41 05 07.7	803±102	0.086	19.33	8.9±0.1	2.3±0.1	0.41±0.02	614±37	-	-	0.15	0.17	irregular with tidal tail
FN2 003	16 35 25 +40 55 51	16 35 25.22 +40 55 42.1	682±92	0.001	17.45	13.9±0.1	5.5±0.1	0.19±0.03	416±19	-	1.70±0.02	0.17	0.03	Sa
FN2 004	16 34 01 +41 20 49	16 34 01.82 +41 20 52.5	666±91	1.000	16.41	20.4±0.1	9.6±0.1	0.40±0.02	403±26	-	1.8±0.02	0.05	0.03	SBa
FN2 005	16 32 43 +41 08 38	16 32 42.39 +41 08 46.1	522±78	0.011	18.67	6.1±0.2	2.1±0.1	0.16±0.03	399±27	-	0.67±0.04	0.02	0.26	Sa
FN2 007	16 35 45 +40 39 14	16 35 46.91 +40 39 03.4	316±60	0.022	17.95	6.1±0.2	2.4±0.1	0.09±0.03	84±25	-	0.92±0.02	0.02	0.12	compact source with 2 nuclei
FN2 008	16 35 47 +41 28 58	16 35 48.04 +41 28 30.3	293±58	0.084	18.32	4.5±0.1	1.3±0.1	0.12±0.03	74±18	-	0.38±0.01	0.10	0.14	Sb
FN2 010	16 35 38 +41 16 58	16 35 36.16 +41 17 27.3	285±57	0.451	19.47	2.8±0.6	-	0.06±0.03	74.00±17	-	0.80±0.01	0.02	0.17	pair: irregular with tidal tail and Sa
FN2 011	16 38 07 +40 58 12	16 38 08.78 +40 58 07.4	260±55	0.118	19.78	<1.97	-	<0.12	<0.44	<0.26	-	0.51	-	Sa?
FN2 012	16 34 13 +40 56 45	16 34 11.98 +40 56 52.8	249±54	0.088	19.99	2.7±0.1	1.0±0.1	<0.08	83±10	-	0.51±0.01	0.07	0.14	Sa?
FN2 015	16 36 07 +40 55 37	16 36 07.71 +40 55 47.1	223±51	0.015	18.64	5.5±0.1	1.2±0.1	<0.06	109±11	-	0.81±0.02	0.26	0.17	interacting galaxies: Sa/b and Sa
FN2 016	16 34 26 +40 54 07	16 34 23.90 +40 54 10.0	218±51	0.191	19.88	3.1±0.1	1.0±0.1	0.06±0.03	<0.25	0.20±0.08	0.55±0.01	0.10	0.13	edge-on: type not clear
FN2 017	16 34 44 +41 08 42	16 34 44.90 +41 08 20.6	213±50	0.030	18.46	2.2±0.7	-	<0.10	<0.25	<0.18	-	0.18	-	interacting galaxies: Sa with edge-on
FN2 018	16 33 38 +41 01 15	16 33 37.23 +41 01 09.1	212±50	0.037	19.51	<2.98	-	0.08±0.03	<0.28	<0.19	-	0.15	-	S0/a
FN2 019	16 37 17 +40 48 36	16 37 16.80 +40 48 25.6	205±49	0.001	18.05	<3.99	-	0.06±0.04	<0.26	<0.18	-	0.10	0.03	Sb
FN2 020	16 32 41 +41 06 10	16 32 40.50 +41 06 15.4	201±49	0.086	20.21	<6.58	-	<0.07	<0.25	<0.11	-	0.15	-	-
FN2 021	16 37 58 +40 51 21	16 37 59.39 +40 51 15.8	196±49	0.001	17.57	<2.18	-	<0.07	<0.19	<0.17	-	0.10	-	-
FN2 022	16 37 08 +41 28 26	16 37 08.21 +41 28 56.1	190±48	0.140	18.80	2.5±0.1	1.7±0.2	<0.10	<0.28	<0.23	0.56±0.02	0.05	0.17	Sa
FN2 023	16 33 51 +40 49 44	16 33 51.65 +40 49 46.3	188±48	0.018	19.04	2.5±0.7	-	<0.11	<0.28	<0.32	-	0.02	-	Sb/c
FN2 025	16 36 31 +40 47 38	16 36 31.23 +40 47 24.7	184±48	0.119	20.24	2.1±0.7	-	0.11±0.03	<0.33	0.11±0.07	-	0.20	-	Sa
CFN2 029	16 34 20 +41 06 54	16 34 19.46 +41 06 37.8	178±47	0.070	19.62	<4.36	-	0.06±0.02	<0.21	<0.24	-	0.15	-	-
CFN2 030	16 35 23 +40 38 42	16 35 22.81 +40 38 37.1	178±47	0.086	20.26	<3.66	-	0.06±0.03	<0.24	0.17±0.05	-	0.15	-	S0/a
CFN2 034	16 34 12 +40 46 26	16 34 12.52 +40 46 34.4	166±46	0.030	19.42	<2.61	-	<0.08	<0.11	<0.10	-	0.55	-	S0/a
CFN2 036	16 37 01 +40 43 08	16 36 59.92 +40 42 46.6	165±46	0.119	19.40	<5.08	-	0.04±0.03	<0.25	<0.28	-	0.10	-	Sa
CFN2 038	16 34 32 +41 22 37	16 34 31.57 +41 22 45.7	161±45	0.023	19.26	1.8±0.1	-	<0.08	<0.17	0.12±0.05	<0.15	0.02	0.14	Sb/c
CFN2 039	16 36 13 +40 42 25	16 36 13.65 +40 42 30.0	160±45	0.001	17.15	5.9±0.2	2.5±0.1	0.11±0.00004	90±21	-	0.88±0.01	0.02	0.07	Sb
CFN2 044	16 37 26 +40 45 39	16 37 25.93 +40 45 37.1	150±44	0.053	19.96	<2.87	-	0.02±0.05	<0.30	<0.14	-	0.05	-	S0/a
CFN2 047	16 34 51 +41 20 27	16 34 49.54 +41 20 49.2	147±44	0.125	19.16	2.4±0.1	1.3±0.1	<0.05	<0.20	<0.21	0.15±0.04	0.07	0.25	S0/a
CFN2 049	16 37 42 +41 19 11	16 37 41.44 +41 19 14.8	143±44	0.210	20.87	1.9±0.2	-	0.09±0.03	74±24	-	<0.15	0.51	-	Sa?

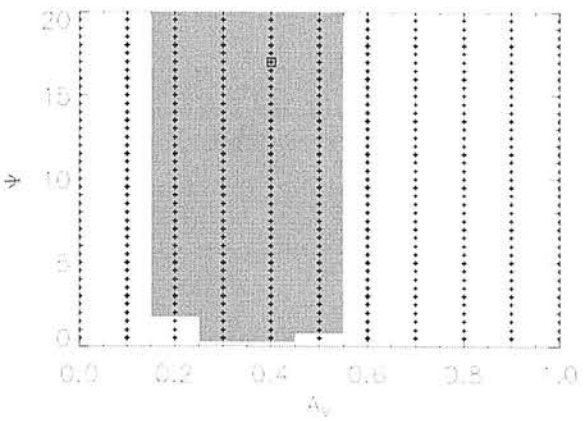

Figure 8. 1σ non–starburst model parameter confidence region for A_V and ψ for FN2 001. Possible models parameter combinations are shown with small crosses (note $A_V = 0 - 10$ and $\psi = 0.5 - 20.0$ were allowed), the best-fitting model is marked with a square and the 1σ confidence region is shaded. **A high resolution version of this figure is available at http://www.roe.ac.uk/~elt/N2paper_images.html .**

model both cirrus and starburst emission or a combination of the two. The models used are briefly described below.

Ultraviolet to submillimetre emission from stars embedded in interstellar dust in galaxies known as the 'infrared cirrus component' are modelled. The model consists of the input stellar radiation field, an interstellar dust model and the radiative transfer treatment of the interaction between the two which generates the output SED.

The input stellar radiation field is taken from the Galaxy Isochrone Synthesis Spectral Evolution Library (GISSEL, Bruzual and Charlot 1993). This gives the radiation from a mass of stars from the ultraviolet to near-infrared as a function of time.

Stars form in molecular clouds and at the early stages of their lives remain inside them during which time their radiation is absorbed by the dust and reprocessed to the infrared. During this phase of their life the model uses the code of Efstathiou et al. (2000) to compute the radiation emitted by stars and dust. The cloud is assumed to disperse 7.2×10^7 years after star formation. Before this, however, due to non-spherical evolution of the cloud, a fraction of the starlight f may have been able to escape without dust absorption. This occurs a time t_m after star formation.

The radiation field within the galaxy is modelled to be due to a large number of randomly oriented molecular clouds whose average emission, for stars in the age range t_m to 7.2×10^7 years, is approximately

$$(1-f)S_\nu^S + fS_\nu^*$$

were S_ν^S is the emission from a giant spherical molecular cloud (Efstathiou et al. 2000) and S_ν^* is the emission from the stellar population (Bruzual and Charlot 1993). The emission of stars younger than t_m is simply S_ν^S and those older than 7.2×10^7 years S_ν^*. The fraction of general starlight that escapes the galaxy unattenuated is also parameterised in a 'leak' variable.

The dust temperature is determined by the intensity of the stellar radiation field and is characterised in the model by the ratio of the bolometric intensity of the radiation field to that in the stellar radiation field in the solar neighbourhood (Mathis et al. 1983) , ψ. The star formation rate, $\dot{\phi}_*(t)$, is assumed to have an exponential form with a time scale, τ, and a Salpeter IMF from $0.1 - 125$ $M_\odot$.

$$\dot{\phi}_*(t) \propto e^{\frac{-t}{\tau}}$$

In the case where a starburst is included the star formation rate is modified to

$$\dot{\phi}_*^{SB}(t) \propto \dot{\phi}_*(t) + f_{SB} e^{\frac{-(t-t_{SB})}{\tau_{SB}}}$$

where f_{SB} is the ratio of the star formation rate at the peak of the starburst to that at time 0, t_{SB} is the age of the galaxy at the start of the starburst and τ_{SB} is the exponential timescale of the starburst.

The interstellar dust model of Siebenmorgen and Kruegel (1992) is used, which models the effects of both small grains and PAHs. The opacity of the dust is characterised by A_V, which controls the amount of UV to near-IR light absorbed by dust and re-emitted in the IR and sub–mm and therefore the ratio of the luminosity in these two bands. The model also has the ability to increase the dust extinction of the starburst component of the galaxy separately via the parameter τ_v, values of 50, 100, 150, 200 are suppported. In the starburst case a grid of models was produced over f_{SB}, A_V and τ_v and for non-starburst galaxies a grid of models was produced over A_V and ψ. The values of the parameters used are given in Table 2.

Table 2. Parameter values

	starburst	non-starburst
τ	8 Gyrs	1 Myr
t_{SB}	galaxy age$-t_m$	-
galaxy age	(universe age at z - 0.5 Gyrs)	
τ_{SB}	40 Myrs	-
leak	0.0	0.1
f	1.0	1.0
t_m	20 Myrs	3 Myrs
A_V	0–3	0–10
τ_v	50–200	-
ψ	5	0.5-20.0
f_{SB}	1–100	-

The effects of each of the model parameters on the resulting SED is shown in Figure 4. The solid black line shows an SED produced using the fixed parameters ($\psi = 5$, $\tau = 8 \times 10^9$, $t_m = 20$ Myrs, $\tau_{SB} = 40$ Myrs) and $f_{SB} = 1$, $A_V = 1$, and $\tau_v = 100$ for comparison. A model with increased dust extinction of $A_V = 3$ is shown by the black dotted line: this increases the emission at longer wavelengths and decreases that in the optical, as would be expected. The dash-dot-dot-dot line shows a model with $t_m = 0$ which dramatically changes the shape of the far–infrared peak due to the reduction in the amount of dust reprocessing of radiation from newly formed stars. However, any significant amount of time spent within a molecular cloud produces a far–infrared peak of shape similar to that of the other models shown. The peak of the SED is also increased significantly with an increase in the strength of the starburst, a model with $f_{SB} = 5$ is shown by the grey dashed line. The other param-

eters do not affect the resulting SED vastly, hence, there are degeneracies and a spread of parameter values may result in a very similar SED. This is evidenced by models with $\psi = 10$ (dashed line), $\tau = 8 \times 10^{10}$ (dash-dot), $\tau_{SB} = 80$ Myrs (long dashes) and $\tau_v = 200$ (solid grey line) all having very similar shaped SEDs. The parameter values chosen were in the mid-range of physically viable values as found by Efstathiou et al. (2000). The τ_v range covers values known to fit normal starbursts ($\tau_v = 50$) (Efstathiou et al. 2000) to heavily obscured starbursts such as Arp 220 ($\tau_v = 200$), this range was also used by Farrah et al. (2003) to fit a number of ULIRGs.

After fitting the models to the multiwavelength data, the parameter values for models within the 1σ χ^2 confidence limit of the best-fitting model were examined. The 1σ limit for starburst model fitted SEDs typically encompasses $\tau_v \pm 50$, $A_V \pm 0.5$ and $f_{SB} \pm 1$. There is also a correlation between f_{SB} and A_V such that models with higher f_{SB} values require low A_V values to fit the data. This is due to an increase in each of the parameters increasing the height of the far-infrared peak (as shown in Figure 4), therefore when both the parameters have high values the height of the peak is over-estimated. We also note that it is difficult to constrain τ_v since we do not have data in the region of the spectrum which it most affects. For non-starburst fitted SEDs the confidence region includes $A_V \pm 0.4$ and $\psi = 0-20$. We also note that since the effect of ψ on the shape of the model SEDs is not marked, it is difficult to constrain this parameter and the 1σ confidence region encompasses the whole allowed parameter range. However, since it has little effect on the shape of the model SEDs this will not affect the property estimates calculated from the models. Examples of these confidence regions for both the starburst (FN2 008) and non-starburst (FN2 001) galaxies are shown in Figures 5 to 8.

5 SPECTRAL ENERGY DISTRIBUTIONS

We have compared the observed spectral energy distributions (SEDs) with predictions from the radiative transfer models of Efstathiou and Rowan-Robinson (2003) and Efstathiou et al. (2000) described in Section 4.

Spectroscopic redshifts were used in the fitting of the model SEDs, where available from the ELAIS catalogue (Rowan-Robinson et al. 2004) or elsewhere. In other cases photometric redshifts were drawn from the ELAIS catalogue, and calculated bespoke by ImpZ (Babbedge et al. 2004) for the remainder. We note that since photometric redshifts are typically accurate to about 10 per cent in $(1+z)$, for the spectroscopic redshifts we have available this implies an accuracy of around 0.11 to 0.14. Therefore, given that the solutions are based on a maximum of four optical bands, the agreement between our photometric and spectroscopic redshifts is reasonable, the largest margin of error being about 20 per cent in $(1+z)$. It is also important to note that there are no catastrophic outliers which can occur due to degeneracies in the solution space of photometric codes. These small redshift errors do not affect the fitted SEDs significantly and the resulting errors on derived properties are much smaller than those incurred by the 1σ error on the fitted SED χ^2 value which are given for all derived quantities.

Table 3. Parameters for the starburst radiative transfer model SEDs fitted to the N2 175 μm ISO source SEDs.

FIRBACK	A_V	f_{SB}	τ_v
FN2 002	0.4	2.0	150
FN2 003	0.3	1.0	200
FN2 004	0.0	1.0	200
FN2 005	0.0	2.0	100
FN2 007	0.2	2.0	200
FN2 008	0.0	2.0	150
FN2 010	0.3	3.0	200
FN2 011	0.5	3.0	200
FN2 012	0.0	3.0	100
FN2 016	0.3	2.0	200
FN2 017	0.4	1.0	200
FN2 018	0.0	2.0	200
FN2 019	0.0	1.0	200
FN2 021	0.0	1.0	150
FN2 022	0.1	1.0	150
FN2 025	0.8	3.0	150
CFN2 029	1.1	10.0	200
CFN2 030	0.9	2.0	200
CFN2 034	1.0	1.0	50
CFN2 038	0.0	1.0	150
CFN2 039	0.0	1.0	100
CFN2 044	0.0	4.0	150
CFN2 047	0.0	1.0	50
CFN2 049	0.1	5.0	50

Table 4. Parameters for the quiescent radiative transfer model SEDs fitted to the N2 175 μm ISO source SEDs.

FIRBACK	A_V	ψ
FN2 000	10.0	20.0
FN2 001	0.4	17.0
FN2 015	1.1	20.0
FN2 020	3.7	0.5
FN2 023	0.2	0.5
CFN2 036	1.0	0.5

The N2 175 μm ISO source SEDs are shown in Figures 9 and 10. They appear to fall into two population categories, the first being star bursting galaxies with their starburst region at a high optical depth (fitted model parameters shown in Table 3 and SEDs plotted in Figure 9) and the second dust obscured galaxies with quiescent star formation (fitted model parameters shown in Table 4 and SEDs plotted in Figure 10).

6 STAR FORMATION RATES AND FAR-INFRARED LUMINOSITIES

Star formation rates for the sources were calculated using both an optical U-band estimate (Cram et al. 1998) and a FIR estimate (Kennicutt 1998), given in equations 2 and 3.

$$\mathrm{SFR}_U(\mathrm{M} \geq 5\ \mathrm{M}_\odot) = \frac{\mathrm{L}_U}{1.5 \times 10^{22}\ \mathrm{W\ Hz^{-1}}} 100\ \mathrm{M}_\odot \mathrm{yr}^{-1} \quad (2)$$

$$\mathrm{SFR}_{\mathrm{FIR}} = 4.5 \times 10^{-44} \mathrm{L}_{\mathrm{FIR}}(\mathrm{ergs\ s^{-1}})\ \mathrm{M}_\odot \mathrm{yr}^{-1} \quad (3)$$

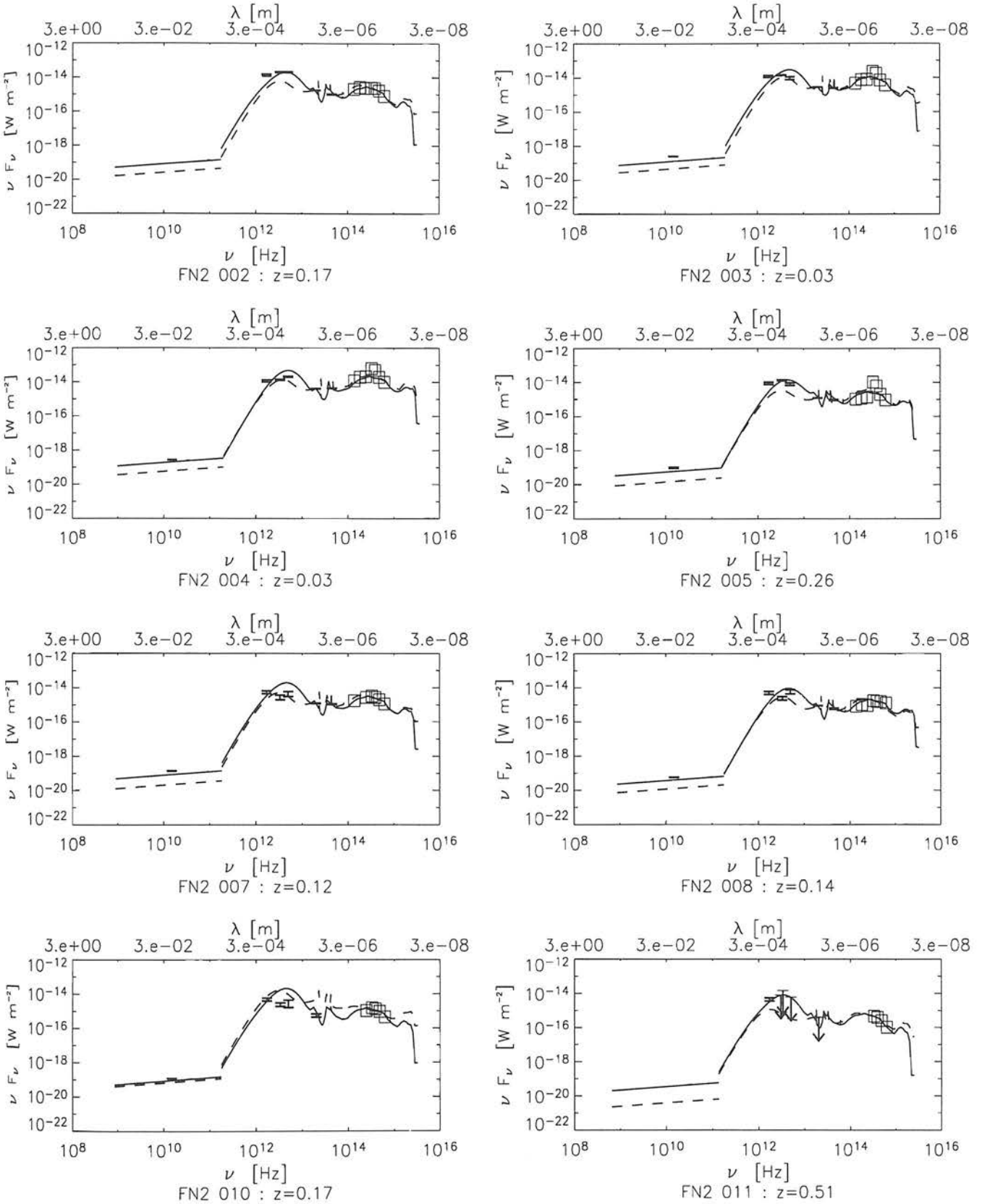

Figure 9. N2 ISO 175 μm sources fitted model SEDs. Best-fitting starburst models (solid line) and cirrus models (dashed line) are shown for each source.

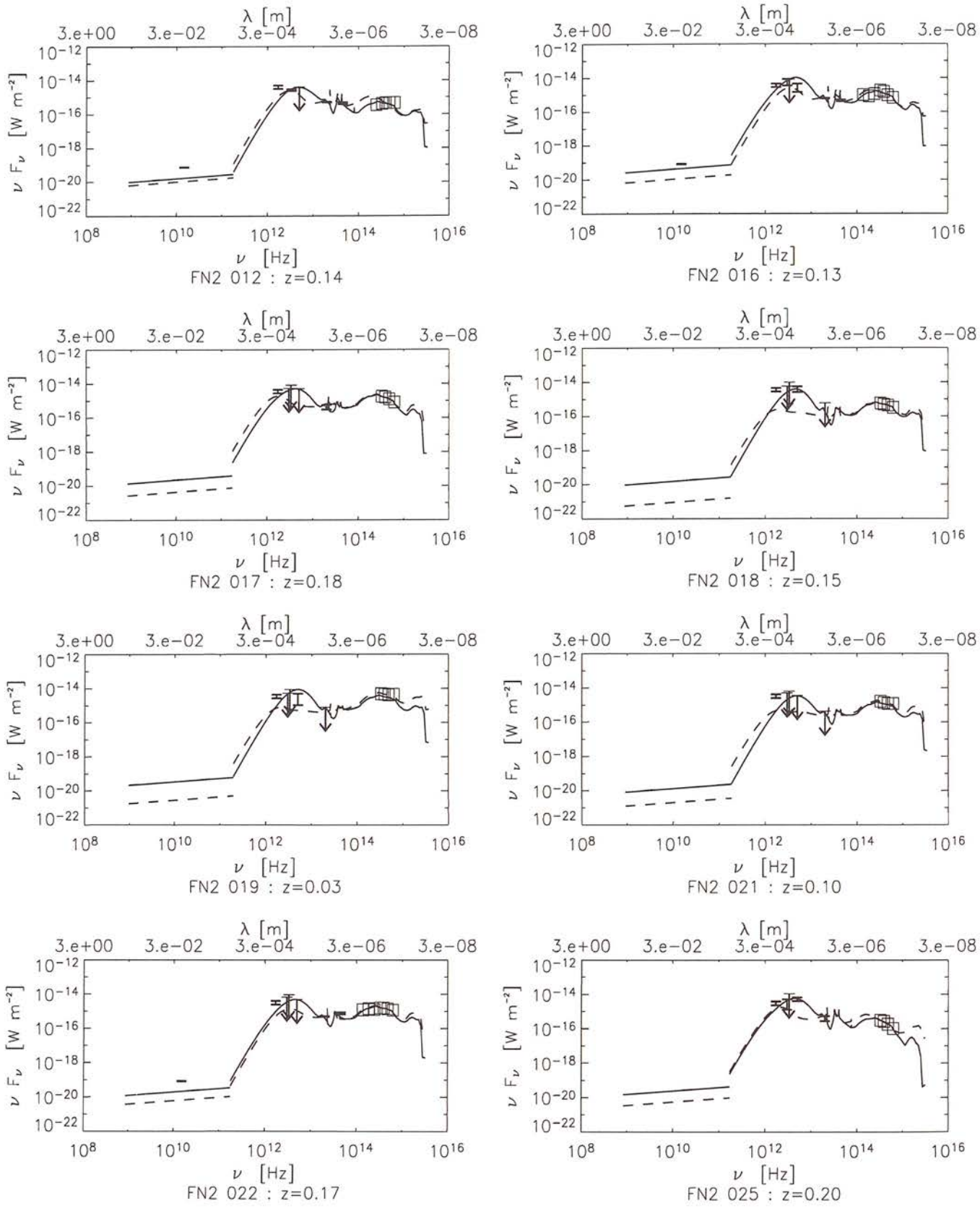

Figure 9 – *continued* N2 ISO 175 μm sources fitted model SEDs. Starburst models (solid line) and cirrus models (dashed line) are shown.

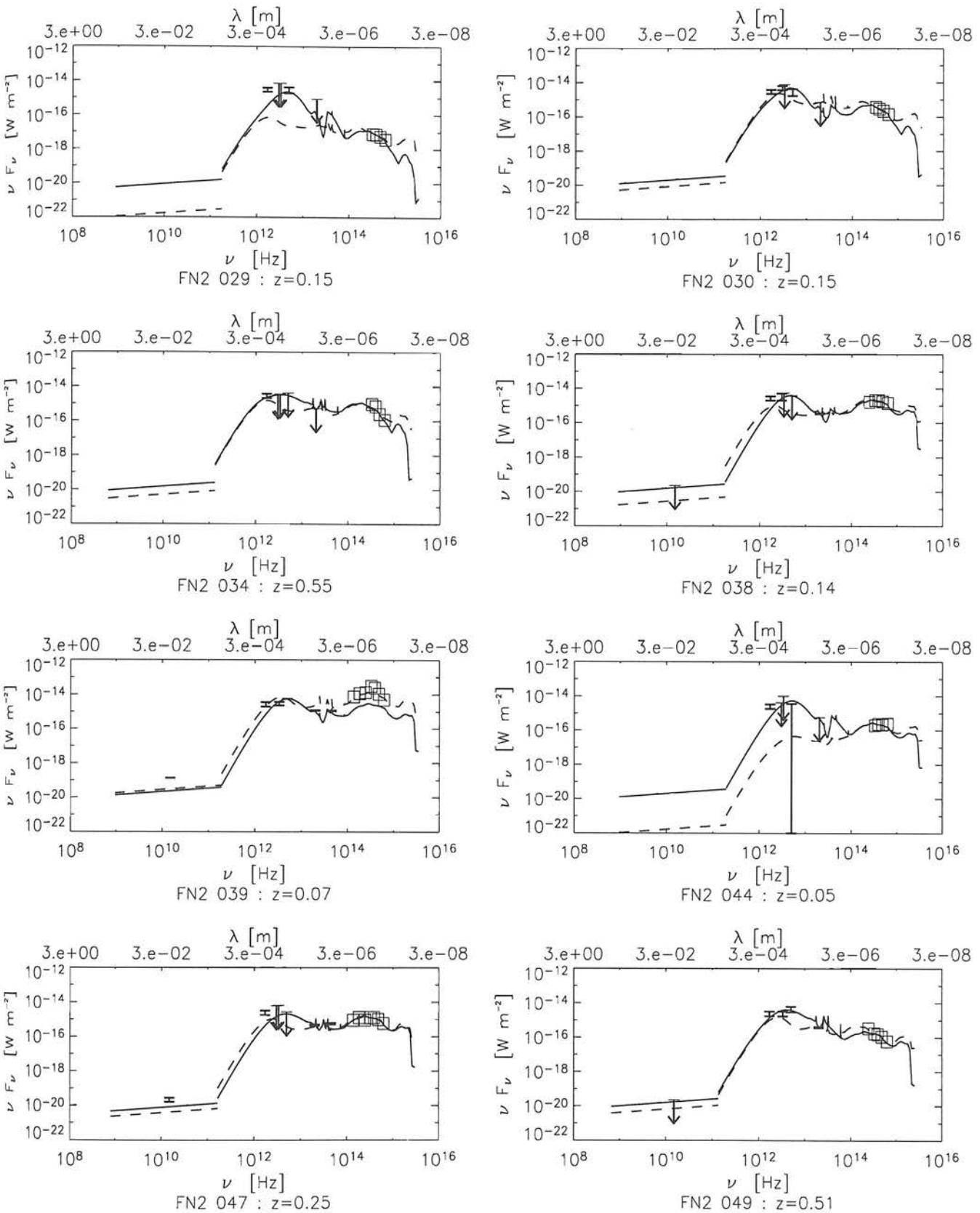

Figure 9 – *continued* N2 ISO 175 μm sources fitted model SEDs. Starburst models (solid line) and cirrus models (dashed line) are shown.

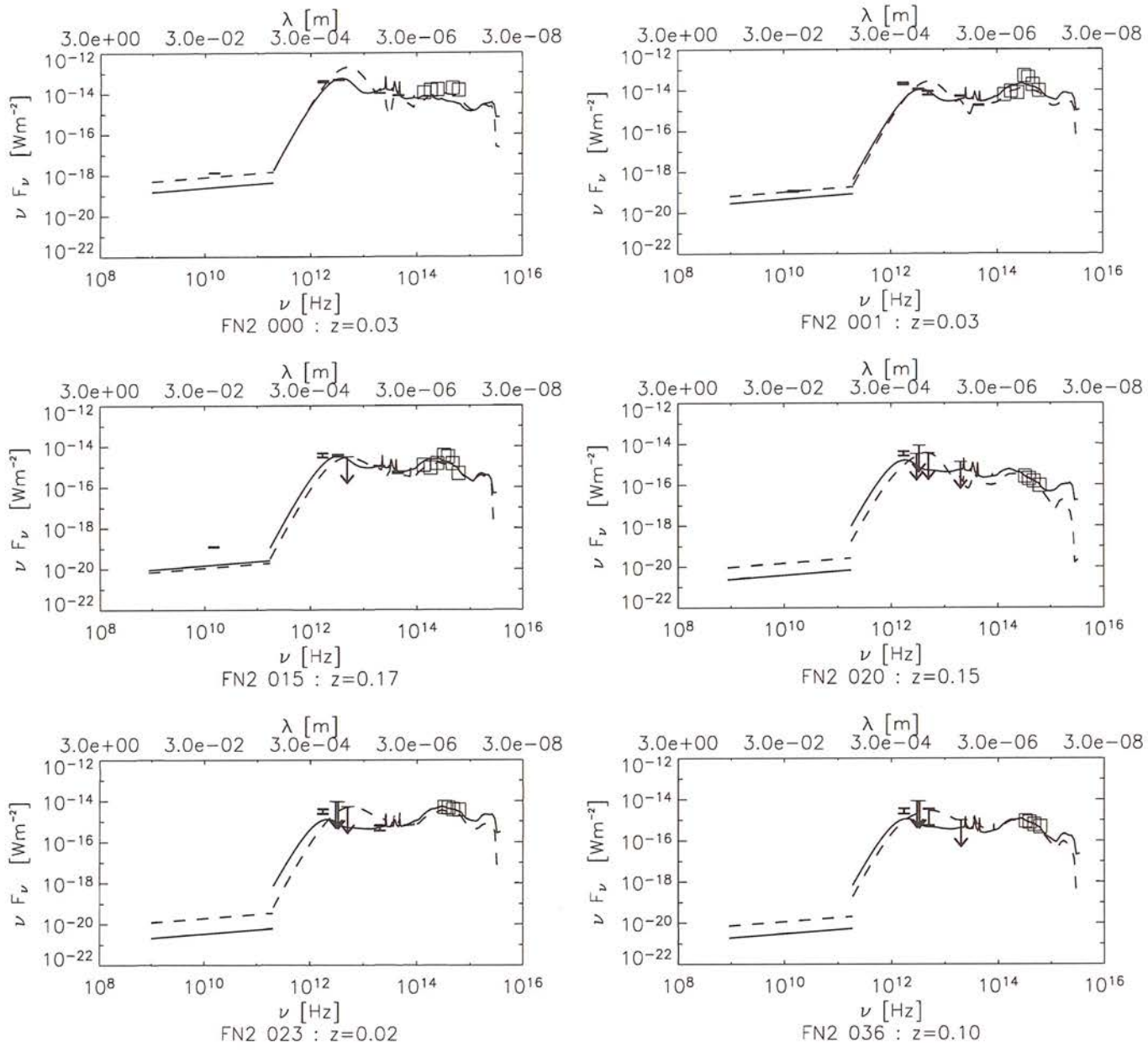

Figure 10. N2 ISO 175 μm sources fitted model SEDs. Cirrus models (solid line) and starburst models (dashed line) are shown.

Where L_U is the U-band luminosity, estimated through interpolation of the SED models and L_{FIR} is the integrated infrared luminosity over 8–1000 μm. These both assume a Salpeter IMF. The chosen IMF affects the estimated star formation rates since we integrate over only a portion of the luminosity of the galaxy's constituent stars. The proportion of stars emitting with luminosities in the range in question will depend on the initial mass function, since it describes the initial distribution of stellar luminosities over the range, as discussed by Mann et al. (2002). The star formation rates for individual sources can be seen in Table 5. The distribution of FIR star formation rates and a comparison between those estimated using both the FIR and U-band luminosities can be seen in Figure 11. There is a large spread in star formation rates with a number of galaxies with SEDs suggestive of a modest starburst. This implies that these galaxies are of low mass since their existing stellar population is dominated by the small starburst at infrared wavelengths. Using star formation rates estimated for SED models falling in the 1σ χ^2 range (see Section 4) the U-band and FIR SFRs are estimated to be accurate to 25 per cent and 50 per cent respectively. It can be seen in the lower pannel of Figure 11 that the FIR star formation rate estimates are significantly higher, typically by a factor of 4, than those using the U-band estimate. This is to be expected for such highly obscured objects.

The integrated far–infrared luminosities over the wavelength range 8-1000 μm were estimated. Results can be seen in Table 5 and a distribution in Figure 12. These are estimated to be accurate to approximately 25 per cent.

Table 5. Star formation rates and FIR luminosities of the starburst galaxies (first block) and cirrus galaxies (second block).

FIRBACK	U–band SFR $M_\odot yr^{-1}$	FIR SFR $M_\odot yr^{-1}$	L_{FIR} $L_\odot$	z
FN2 002	36.9	93.1	4.87×10^{11}	0.17
FN2 003	3.8	3.6	1.83×10^{10}	0.03
FN2 004	9.4	5.4	2.85×10^{10}	0.03
FN2 005	131.4	169.1	9.18×10^{11}	0.26
FN2 007	22.3	39.5	2.07×10^{11}	0.12
FN2 008	24.4	27.3	1.46×10^{11}	0.14
FN2 010	21.7	88.9	4.68×10^{11}	0.17
FN2 011	104.8	391.9	2.04×10^{12}	0.51
FN2 012	6.3	13.1	7.07×10^{10}	0.14
FN2 016	12.6	25.6	1.33×10^{11}	0.13
FN2 017	32.5	28.1	1.41×10^{11}	0.18
FN2 018	8.4	12.2	6.53×10^{10}	0.15
FN2 019	1.7	1.0	5.22×10^{9}	0.03
FN2 021	7.9	4.8	2.59×10^{10}	0.10
FN2 022	32.6	22.3	1.18×10^{11}	0.17
FN2 025	5.8	39.7	2.07×10^{11}	0.20
CFN2 029	0.1	7.2	3.79×10^{10}	0.15
CFN2 030	3.2	16.8	8.60×10^{10}	0.15
CFN2 034	138.2	296.6	1.48×10^{12}	0.55
CFN2 038	25.6	11.6	6.19×10^{10}	0.14
CFN2 039	7.7	4.1	2.23×10^{10}	0.07
CFN2 044	0.4	1.8	9.41×10^{9}	0.05
CFN2 047	63.2	27.2	1.49×10^{11}	0.25
CFN2 049	49.1	252.8	1.39×10^{12}	0.51
FN2 000	2.8	7.9	3.97×10^{10}	0.03
FN2 001	9.3	1.7	8.37×10^{9}	0.03
FN2 015	29.1	18.8	9.58×10^{10}	0.17
FN2 020	3.0	7.5	2.56×10^{10}	0.15
FN2 023	1.1	0.1	3.87×10^{8}	0.02
CFN2 036	4.3	2.5	8.82×10^{9}	0.10

Table 6. Greybody parameters and dust masses for the starburst galaxies (first block) and cirrus galaxies (second block).

FIRBACK	T (K)	β	$M_{dust}(M_\odot)$
FN2 002	40	1.4	8.0×10^{7}
FN2 003	36	1.5	4.3×10^{6}
FN2 004	41	1.9	1.4×10^{6}
FN2 005	40	1.9	2.9×10^{7}
FN2 007	42	1.5	2.5×10^{7}
FN2 008	39	1.9	6.5×10^{6}
FN2 010	45	1.4	5.4×10^{7}
FN2 011	41	1.4	2.5×10^{8}
FN2 012	39	1.9	3.0×10^{6}
FN2 016	42	1.4	2.1×10^{7}
FN2 017	36	1.4	3.9×10^{7}
FN2 018	39	2.0	2.4×10^{6}
FN2 019	41	1.9	2.5×10^{5}
FN2 021	41	1.9	9.9×10^{5}
FN2 022	41	1.6	1.0×10^{7}
FN2 025	40	1.3	4.5×10^{7}
CFN2 029	49	1.2	5.6×10^{6}
CFN2 030	38	1.3	2.5×10^{7}
CFN2 034	35	1.1	6.5×10^{8}
CFN2 038	40	1.9	2.6×10^{6}
CFN2 039	40	1.9	9.2×10^{5}
CFN2 044	41	1.9	4.2×10^{5}
CFN2 047	41	1.7	6.6×10^{6}
CFN2 049	44	1.5	6.6×10^{7}
FN2 000	31	1.9	7.2×10^{6}
FN2 001	30	1.9	1.7×10^{6}
FN2 015	31	1.9	1.4×10^{7}
FN2 020	17	1.8	1.6×10^{8}
FN2 023	16	2.0	2.0×10^{6}
CFN2 036	17	1.8	5.1×10^{7}

7 DUST TEMPERATURES AND MASSES

The dust in the galaxy radiative transfer models is at a range of temperatures, however, we computed an emission-weighted estimate via the fitting of a grey body to the far-infrared portion of the SED (from the peak of the SED to 1000 μm) as follows:.

$$F_\nu \propto \frac{\nu^{\beta+3}}{e^{\frac{h\nu}{kT}} - 1} \quad (4)$$

In equation 4 T is the dust temperature which controls the position of the maximum of the long wavelength SED, β is the frequency dependence of the dust grain emissivity and controls the gradient of the long wavelength slope of the grey body curve. Fitting over this wavelength range allows us to constrain both the temperature and β, without reaching the far-infrared where the dust may go optically thick and no longer be well described by a greybody model. The derived values for the sources can be seen in Table 6. Using the method described above (Section 6) for the starburst models these are estimated to be accurate to T$\pm$3 K and $\beta \pm 0.5$, in the non-starburst case these estimates are accurate to T$\pm$14 K and $\beta \pm 0.1$.

The distribution of the dust temperatures for the N2 175 μm ISO source galaxies is shown in Figure 13. The starburst galaxy dust temperatures are slightly higher than those found by Sajina et al. (2003) ($\sim 20-30$ K) for a sample of N1 ISO 175 μm sources but have similar emissivity coefficients: Sajina et al. (2003) found $\beta \sim 1.5 - 1.7$.

Dust masses were estimated using the submillimetre prescription (Hildebrand 1983) at a rest wavelength of 450 μm:

$$M_{dust} = \frac{S_{\nu_0} D_L^2}{(1+z)\kappa(\nu_r) B(\nu_r, T_{dust})} \quad (5)$$

$$\kappa(\nu_r) = 0.067 \left(\frac{\nu_r}{2.5 \times 10^{11}}\right)^\beta \quad (6)$$

where ν_0 and ν_r are the observed and rest frame frequencies respectively, S_{ν_0} is the observed flux at ν_0, estimated through interpolation of the SED models, and $B(\nu_r, T_{dust})$ is the Planck function in the rest frame. T_{dust} and β are the dust temperature and dust grain emissivity estimated via the greybody fitting. $\kappa(\nu_r)$ is known as the mass absorption coefficient and has units of $m^2 kg^{-1}$. This approach assumes that the galaxy is optically thin at submillimetre wavelengths. Resulting dust masses, calculated at a wavelength of 450 μm (accurate to approximately 25 per cent), are shown in Table 6. We note that the dust masses follow the trend of the far–infrared luminosities and star formation rates, those with high SFRs and L_{FIR} having larger dust masses.

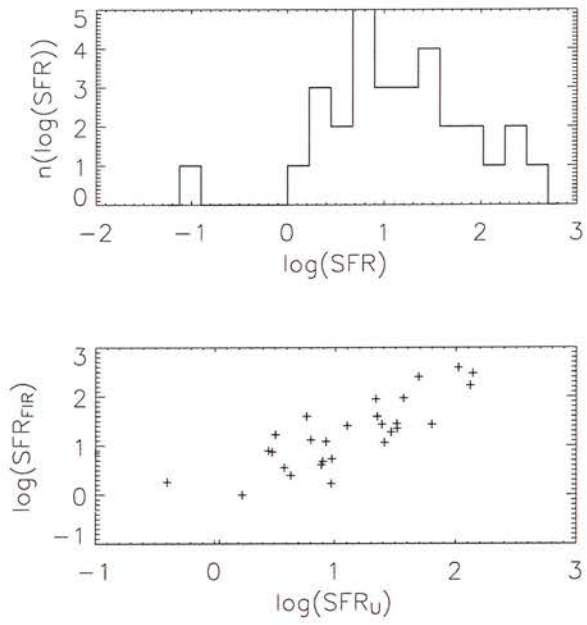

Figure 11. Top window:Far–infrared star formation rate distribution ($M_\odot yr^{-1}$) for the ISO 175 μm sources, see Table 5. Bottom window: Far–infrared starformation rate estimates vs. u-band starformation rate estimates.

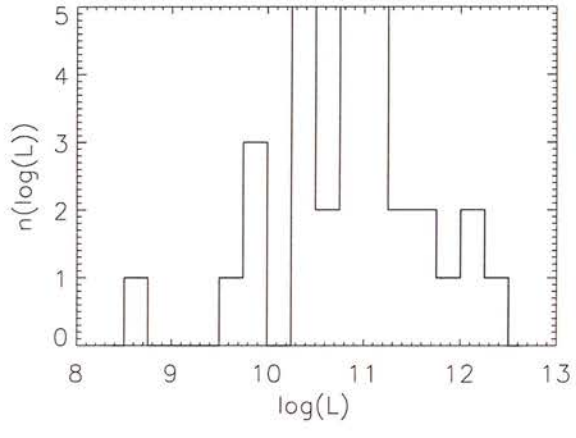

Figure 12. Distribution of far–infrared luminosities ($L_\odot$) for the ISO 175 μm sources, see Table 5.

8 EVOLUTIONARY MODELS

The detection of the far–infrared background in 1996 led to a re–evaluation of galaxy evolution models. One of the earliest new models to emerge was that of Burigana et al. (1997) which used the FIRB to place constraints on models for the cosmic star formation history and therefore the production of the background signal. This paper concluded that stellar nucleosynthesis is able to account for a FIRB intensity not far from the upper limit set by the COBE detection. It also suggested substantially higher star formation rates at higher redshifts than suggested by optical data.

It is now fairly well established that infrared–selected

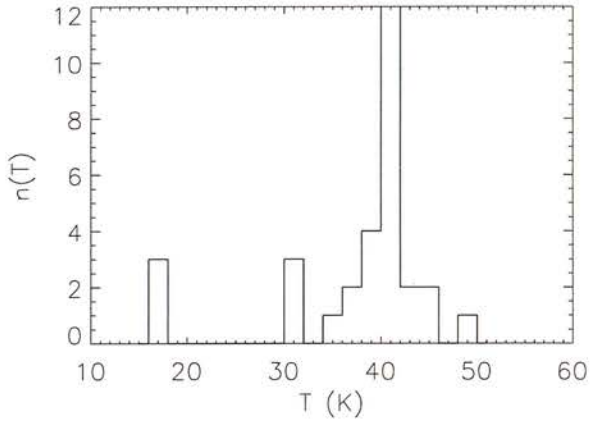

Figure 13. Distribution of dust temperatures for the ISO 175 μm sources, see Table 6.

galaxies undergo extremely high rates of evolution compared to measurements at other wavelengths over the approximate redshift range $0 < z < 1$. Chary and Elbaz (2001) models imply that about 80 per cent of the cosmic infrared background was produced between $0 < z < 1.5$ compared to only 30 per cent of the 850 μm background. Also the dust enshrouded star formation rate peaks at $z \sim 0.8$ at a value of ~ 0.25 $M_\odot yr^{-1} Mpc^{-3}$ and at least 70 per cent of this star formation takes place in infrared luminous galaxies with $L_{IR} > 10^{11}$ $L_\odot$. In fact ~ 75 per cent of 15 μm sources examined by Elbaz et al. (2002) were found to be in this category of luminous infrared galaxies and have star formation rates of ~ 100 $M_\odot yr^{-1}$. The comoving density of infrared light due to these galaxies was over 40 times larger at $z \sim 1$ than it is today.

Models of the cosmic infrared background produced by Lagache et al. (2003), that are able to fit number counts and unresolved anisotropies imply that the luminosity function must change dramatically with redshift. These also show rapid evolution of high luminosity sources up to z=1. It has been suggested that this strong evolution is due to bimodal star formation, one long lived quiescent, passive phase and one enhanced starburst phase triggered by interactions and mergers (e.g. Lagache et al. 2003, Franceschini et al. 2001). Franceschini et al. (2001) also suggest that the interpretation of evolution with redshift could be biased due to an increase in the probability of detecting a galaxy during an active phase caused by both an increase of the rate of interactions, in part due to the geometry of the expanding universe, and an increase of the infrared luminosity due to more abundant fuel available in the past.

However, Chapman et al. (2002) point out that at $z < 1$ the coldest and most dusty galaxies will have the greatest flux densities for an equivalent infrared luminosity. At $z > 1$ the FIRBACK 175 μm detection will lie bluewards of the peak in the restframe black body spectrum and dust temperatures of < 100 K will no longer have significant effects on the observed flux. Therefore at higher redshifts, the flux limited FIRBACK survey will be biased towards cooler galaxies.

Models such as Franceschini et al. (2001) and

Chary and Elbaz (2001) do not explicitly incorporate such cold luminous galaxies which are preferentially selected in our survey for $z < 1$. Chapman et al. (2002) examined two of the 175 μm selected ISO sources and indentified them as unusually cold ULIRGs with dust temperatures of ~ 30 K which is significantly lower than the usual ULIRG dust temperature of ~ 50 K. Cold, dusty yet luminous objects suggest large masses of dust heated at a moderate intensity. The existence of such sources may seriously affect the interpretation of SCUBA sources if they are numerous at high-z. Since many models rapidly evolve a hot ULIRG population to the local luminosity function they may predict a false redshift distribution if these cold sources are not considered.

9 DISCUSSION AND CONCLUSIONS

We have found bright galaxy optical identifications for 31 out of the 55 N2 ISO 175 μm sources and stellar identifications for 2 sources. 17 of those with galaxy optical IDs were associated with sources from the band-mergerd ELAIS catalogue. We then compard the spectral energy distributions of the sources with predictions from the radiative transfer models of Efstathiou and Rowan-Robinson (2003) and Efstathiou et al. (2000).

22 of the 175 μm sources have not been confidently associated with optical counterparts and therefore their nature is still unknown. However, only 4 of these sources are in the main FIRBACK catalogue, the remaining 18 are listed in the complementary catalogue which contains detections at lower signal to noise and therefore they may just be spurious detections. However, these could also be due to sources having multiple IDs or them being associated with heavily extincted objects which were not in the optical catalogue.

24 galaxies have been classed as star bursting. They have a range of far-infrared luminosities from 10^9 $L_\odot$ to 10^{12} $L_\odot$ corresponding to a range of star formation rates from ~ 0.1 to ~ 400 $M_\odot yr^{-1}$. Although those at the lower end of this range are not forming stars at a dramatic rate their small starburst component is dominating over the existing population at far-infrared wavelengths. This implies that the 175 μm FIRBACK sample contains a significant number of low mass galaxies. The majority of these galaxies are spirals and a number are S0 galaxies.

Sources FN2 005, 011, CFN2 034 and CFN2 049 have $L_{FIR} \sim 10^{12}$ $L_\odot$ and high star formation rates, we therefore suggest that these are in the low redshift tail of the rapidly evolving ULIRG population. This is in agreement with the findings of Sajina et al. (2003) who found one sixth of their 175 μm sample to fall into this category. We also suggest that those galaxies with moderate star formation rates and far-infrared luminosities of $\sim 10^{11}$ $L_\odot$ at moderate z (≥ 0.12) are LIRGs (FN2 002, 007, 008, 010, 016, 017, 022, 025 and CFN2 047). Those with $z \leq 0.12$ are more normal galaxies perhaps with lower masses (FN2 003, 004, 012, 018, 019, 021, CFN2 029, 030, 038, 039, 044).

The estimated dust temperatures of the starburst galaxies are 30-49 K which is relatively low compared to that of standard ULIRGs which have $T_D \sim 50$ K. The most luminous of these galaxies also have large dust masses adding weight to the suggestion that cold, dusty, luminous objects have large masses of dust heated at a moderate intensity (Chapman et al. 2002).

A further 6 sources have non-starburst SEDs. These sources have far-infrared luminosities of $10^8 - 10^{10}$ $L_\odot$ and star formation rates of < 30 $M_\odot yr^{-1}$. We suggest that these are quiescently star forming galaxies (FN2 000, 001, 015, 020, 023, 036). These galaxies have low dust temperatures of $16-31$ K and dust masses of $10^6 - 10^8$ $M_\odot$. 3 of these galaxies are spirals, 1 is an intereacting galaxy pair, 1 is either an irregular or a spiral with tidal tail and the morphology of the remaining source was unable to be deduced.

Lagache et al. (2003) predicted 62 per cent of 175 μm sources would have $z < 0.25$ and the remainder $0.8 < z < 1.2$. We find only one 175 μm source with $S > 180$ mJy with $z > 0.25$ and therefore have $\gg 62$ per cent with $z < 0.25$. However, our mix of higher redshift ULIRGs and lower redshift starburst galaxies is in good qualitative agreement with the predictions of the Lagache et al. (2003) starburst templates as adopted by Sajina et al. (2003). Lagache et al. (2003) also state that to gain agreement between their models and observations they require a cold local population which is what has been found.

There has also been recent investigation into the submillimetre properties of the local universe through the SCUBA Local Universe Galaxy Survey (Dunne and Eales 2001; Dunne et al. 2000). This survey found a population of galaxies with $L_{FIR} = 10^{10} - 10^{11}$ $L_\odot$ and dust masses of $10^7 - 10^8$ $M_\odot$. The population was fitted with a two-component dust model to discover cold dust temperatures of ~ 21 K and hot dust temperatures of $30-50$ K. It is clear that the local 175 μm FIRBACK population has very similar properties to these galaxies. The sources of the far-infrared background are therefore offering a stepping stone between high redshift SCUBA galaxies and the local universe.

10 ACKNOWLEDGEMENTS

E. L. Taylor acknowledges support from a PPARC studentship. Ph. Héraudeau acknowledges support from the EU TMR Network "SISCO" (HPRN-CT-2002-00316). S. Mei thanks Marc Postman for insights on the galaxy morphological classification. We thank an anonymous referee for a thorough reading of our paper which significantly improved the presentation of its results.

REFERENCES

Babbedge T. S. R. et al., 2004, MNRAS, 353, 654
Blain, A. W., Smail, I., Ivison, R. J., Kneib, J.-P., 1999, MNRAS, 302, 632
Bruzual A., G. and Charlot, S., 1993, ApJ, 405, 538
Burigana, C., Danese, L., de Zotti, G., Franceschini, A., Mazzei, P., Toffolatti, L., 1997, MNRAS, 287, L17
Chapman, S. C., Smail, I., Ivison, R. J., Helou, G., Dale, D. A., and Lagache, G., 2002, ApJ, 573, 66
Chary, R. and Elbaz, D., 2001, ApJ, 556, 562
Ciliegi, P. et al., 1999, MNRAS, 302, 222
Cram, L., Hopkins, A., Mobasher, B., and Rowan-Robinson, M., 1998, ApJ, 507, 155
Dennefeld, et al., 2005, A&A, submitted

Dole, H. et al., 1999, usis.conf.1031D
Dole, H. et al., 2001, A&A, 372, 364
Dunne, L., Eales, S., Edmunds, M., Ivison, R., Alexander, P., and Clements, D. L., 2000, MNRAS, 315, 115
Dunne, L. and Eales, S. A., 2001, MNRAS, 327, 697
Efstathiou, A. and Rowan-Robinson, M., 2003, MNRAS, 343, 322
Efstathiou, A., Rowan-Robinson, M., and Siebenmorgen, R., 2000, MNRAS, 313, 734
Elbaz, D., Cesarsky, C. J., Chanial, P., Aussel, H., Franceschini, A., Fadda, D., and Chary, R. R., 2002, A&A, 384, 848
Farrah, D., Afonso, J., Efstathiou, A., Rowan-Robinson, M., Fox, M., Clements, D., 2003, MNRAS, 343, 585
Franceschini, A., Aussel, H., Cesarsky, C. J., Elbaz, D., and Fadda, D., 2001, A&A, 378, 1
González–Solares E. et. al., 2003, MNRAS, submitted astro-ph/0402406
Greve, T. R., Ivison, R. J., Bertoldi, F., Stevens, J. A., Dunlop, J. S., Lutz, D., Carilli, C. L., 2004, MNRAS, 354, 779
Hacking, P. B. and Soifer, B. T., 1991, ApJ, 367, 49
Héraudeau, Ph. et al., 2004, MNRAS, 354, 924
Hildebrand, R. H., 2004, QJRAS, 24, 267
Kennicutt, R. C., 1998, ARA&A, 36, 189
Kessler, M. F. et al., 1996, A&A, 315, L27
Lagache, G., Dole, H., and Puget, J.-L., 2003, MNRAS, 338, 555
Lawrence, A. et al., 1999, MNRAS, 308, 897
Lemke, D. et al., 1996, A&A, 315, 64
Lonsdale, C. J., Hacking, P. B., Conrow, T. P., Rowan-Robinson, M., 1990, ApJ, 358, 60
Madau, P., Pozzetti, L., and Dickinson, M., 1998, ApJ, 498, 106
Mann, R. G. et al., 1997, MNRAS, 289, 482
Mann, R. G. et al., 2002, MNRAS, 332, 549
Mathis, J. S., Mezger, P. G., and Panagia, N., 1983, A&A, 128, 212
McMahon, R. G., Walton, N. A., Irwin, M. J., Lewis, J. R., Bunclark, P. S., Jones, D. H., 2001, NewAR, 45, 97M
Oliver, S. J. and Rowan-Robinson, M. and Saunders, W., 1992, MNRAS, 256, 15
Partridge, R. B., Peebles, P. J. E., 1967, ApJ, 148, 377
Patris, J., Dennefeld, M., Lagache, G., Dole, H., 2003, A&A, 412, 349
Postman, M. et al., 2005, ApJ, submitted
Puget, J.-L., Abergel, A., Bernard, J.-P., Boulanger, F., Burton, W. B., Desert, F.-X., and Hartmann, D., 1996, A&A, 308, L5+
Rowan-Robinson, M. et al., 2004, MNRAS, 351, 1290
Sajina, A., Borys, C., Chapman, S., Dole, H., Halpern, M., Lagache, G., Puget, J.-L., and Scott, D., 2003, MNRAS, 343, 1365
Scott, S. E. et al., 2002, MNRAS, 331, 817
Siebenmorgen, R. and Kruegel, E., 1992, A&A, 259, 614
Silva, L., Granato, G. L., Bressan, A., Danese, L., 1998, ApJ, 509, 103
Soifer, B. T., Neugebauer, G., Houck, J. R., 1987, ARA&A, 25, 187
Sunyaev, R. A., Tinsley, B. M., Meier, D. L., 1978, ComAp, 7, 183
Vaccari, M. et al., 2004, MNRAS, submitted astro-ph/0404315

11 APPENDIX

Notes on the multiwavelength associations for individual sources. Images of each of the sources are shown in Figure 14.

FN2 - 000: Identification with a bright optical galaxy (P_{ran} = 0.026, r' magnitude = 18.5, z_{spec} = 0.03, irregular or edge–on spiral with tidal tail). This is also an ELAIS source (`ELAISC15_J163734.4+405208`) optical ID with detected emission at 15, 6.7, 90 μm and radio detected emission at 20 cm. There are two other ELAIS sources (`ELAISC15_J163729.3+405248`, `ELAISC15_J163731.3+405156`) within the FIRBACK error ellipse with optical IDs with associated P_{ran} values of 0.604 and 0.246.

FN2 - 001: Identification with a bright optical galaxy (P_{ran} = 0.002, r' magnitude = 18.3, z_{phot} = 0.10, edge–on Sb/c). This is also an ELAIS source (`ELAISC15_J163507.7+405929`) optical ID with detected emission at 15, 6.7, 90 μm and radio detected emission at 20 cm. This also has an IRAS detection at 60μm.

FN2 - 002: Identification with an optical galaxy (P_{ran} = 0.078, r' magnitude = 19.3). An ELAIS source ID (`ELAISC15_J163608.1+410507`) has an optical ID with the same galaxy, however, they are different entries in the INT catalogue caused by the splitting up of bright objects by the image analyser. The ELAIS ID has P_{ran} = 0.086, z_{spec} = 0.17, it is an irregulare galaxy with a tidal tail and detected has emission at 15, 6.7 and 90 μm. This also has an IRAS detection at 60 μm. There are two other ELAIS sources (`ELAISC15_J163608.4+410529`, `ELAISC15_J163611.6+410452`) within the FIRBACK error ellipse with optical IDs with associated P_{ran} values of 0.885 and 0.983.

FN2 - 003: Identification with a bright optical galaxy (P_{ran} = 0.001, r' magnitude = 17.5, z_{spec} = 0.03, Sa). This is also an ELAIS source (`ELAISC15_J163525.1+405542`) optical ID with detected emission at 15, 6.7, 90 μm and radio detected emission at 20 cm. This also has an IRAS detection at 60 μm.

FN2 - 004: This ID is a small optical galaxy (P_{ran} = 0.079, r magnitude = 20.2). However, there is also present another large bright galaxy (r magnitude = 16.4) which is mistakenly classified as a star so is forcably not associated and assigned a P_{ran} of 1.0. Previous associations allowing ISO 175 μm sources to be associated with stars chose this ID which matches with the optical ID of an ELAIS source (`ELAISC15_J163401.8+412052`) which has z_{spec} = 0.03 and detected emission at 15, 6.7, 90 μm and radio emission at 20 cm and has the morphology of a SBa. This also has an IRAS detection at 60 μm.

FN2 - 005: Identification with bright optical galaxy (P_{ran} = 0.011, r magnitude = 18.7, z_{spec} = 0.26, Sa). This is also an ELAIS source (`ELAISC15_J163242.4+410847`) optical ID with detected emission at 15, 6.7, 90 μm and radio emission at 20 cm. This also has an IRAS detection at 60 μm. There is another ELAIS source

Figure 14. Optical INT images of the N2 ISO 175 μm sources. FIRBACK source position marked with a diamond (axes show distance from source position in arcsec), optical ID marked with a cross, ELAIS positions marked with small circles, ELAIS optical IDs marked with pluses. A 90 arcsec circle is plotted to show the region in which optical IDs were searched for. **Figure available from http://www.roe.ac.uk/~elt/N2paper_images.html.**

(ELAISC15_J163244.6+410911) within the FIRBACK error ellipse with an optical ID with an associated P_{ran} of 0.315.

FN2 - 006: Identification with what appears on the image to be a pair of interacting galaxies (P_{ran} = 0.001, r' magnitude = 18.2, no redshift). An ELAIS source (ELAISC15_J163506.1+411038) appears to be on the same object, however, is does not have an optical ID. There is another ELAIS source (ELAISC15_J163503.5+411137) within the FIRBACK error ellipse.

FN2 - 007: Identification with bright compact source with two nuclei (P_{ran} = 0.022, r magnitude = 18.0, z_{spec} = 0.12). This is also an ELAIS source (ELAISC15_J163546.9+403903) optical ID with detected emission at 15, 6.7, 90 μm and radio emission at 20 cm. This also has an IRAS detection at 60 μm. There is another ELAIS source (ELAISC15_J163545.5+403825) within the FIRBACK error ellipse with an optical ID with an associated P_{ran} of 0.998.

FN2 - 008: Identification with an optical galaxy (P_{ran} = 0.053, r' magnitude = 18.4). However, the 175 μm flux has been assigned in the ELAIS catalogue to a source (ELAISC15_J163548.0+412829) with a different optical ID. This ID has P_{ran} = 0.084 for association with the FIRBACK source and is therefore a possible identification. The ELAIS source has z_{spec} = 0.14 and detected emission at 15, 6.7, 90 μm and radio emission at 20 cm, it is an Sb galaxy. This also has an IRAS detection at 60 μm.

FN2 - 009: This source does not have a confident optical identification. It is best associated with a galaxy with P_{ran} = 0.440, this does not agree with the optical ID for the ELAIS source (ELAISC15_J163359.2+405303, P_{ran} = 0.880, r' magnitude=18.8) to which the 175 μm flux has been assigned, this source also has detected emission at 15, 6.7, 90 μm and radio emission at 20 cm. This source is confidently identified as a star when stellar associations are allowed with P_{ran} = 0.142 (the threshold for these associations is P_{ran} = 0.4), however, this still does not agree with the ELAIS optical ID.

FN2 - 010: Identification with a pair of interacting galaxies, an irregular with a tidal tail and an Sa (P_{ran} = 0.114, r' magnitude = 18.8). An ELAIS source (ELAISR163536+411726) ID has an optical ID with the same galaxy, however, they are different entries in the INT catalogue caused by the splitting up of bright objects by the image analyser. The ELAIS source has z_{spec} = 0.17 and detected emission at 90 μm and radio emission at 20 cm. This also has an IRAS detection at 60 μm.

FN2 - 011: Identification with an optical galaxy (P_{ran} = 0.118, r' magnitude = 19.8, z_{phot} = 0.51, Sa?). However, the 175 μm flux has been assigned in the ELAIS catalogue to a source (ELAISR163809+405839T) with a different galaxy ID (r magnitude = 18.6) and detected emission only in the far-infrared and radio. This has a P_{ran} value of 0.547 for association with the FIRBACK source, it is therefore not an acceptable association.

FN2 - 012: Identification with an optical galaxy (P_{ran} = 0.006, r' magnitude = 18.5). However, the 175 μm flux has been assigned in the ELAIS catalogue to a source (ELAISC15_J163412.0+405652) with a different galaxy ID (r magnitude = 20.0). This ID has P_{ran} = 0.088 for association with the FIRBACK source and is therefore a possible identification. The ELAIS source has z_{spec} = 0.14 and detected emission at 15, 6.7, 90 μm and radio emission at 20 cm, it is possibly an Sa. There are two other ELAIS sources (ELAISC15_J163417.4+405710, ELAISC15_J163417.9+405653) within the FIRBACK error ellipse with optical IDs with associated P_{ran} values of 0.990 and 1.000.

FN2 - 013: This source does not have a confident optical identification. It is best associated with an optical galaxy (P_{ran} = 0.260, r' magnitude = 20.4). This is also an ELAIS source (ELAISC15_J163406.5+405106) optical ID with detected emission at 15 μm.

FN2 - 014: This source appears in the ELAIS 90 and 175 μm unassociated catalogue (ELAIS-FBK175_N2_014). It does not have a confident optical identification.

FN2 - 015: Identification with a pair of interacting galaxies, an Sa/b and an Sa, (P_{ran} = 0.015, r magnitude = 18.6, z_{spec} = 0.17). This is also an ELAIS source (ELAISC15_J163607.7+405546) optical ID with detected emission at 15, 6.7, 90 μm and radio emission at 20 cm.

FN2 - 016: Identification with an edge–on optical galaxy whose morphological type is unclear (P_{ran} = 0.119, r' magnitude = 20.0). An ELAIS source (ELAISC15_J163423.9+405410) ID has an optical ID with the same galaxy, however, they are different entries in the INT catalogue caused by the splitting up of bright objects by the image analyser. This source has z_{spec} = 0.13 and detected emission at 15, 6.7 μm and radio emission at 20 cm. This also has IRAS detections at 60 and 100 μm. There are two other ELAIS sources (ELAISC7_J163421+405413, ELAISC7_J163423+405506) within the FIRBACK error ellipse one with an optical ID with P_{ran} = 1.000, the other has no optical ID.

FN2 - 017: Identification with a pair of interacting galaxies, an Sa and an edge–on spiral, (P_{ran} = 0.030, r' magnitude = 18.5, z_{phot} = 0.18). The 175 μm flux is assigned to an ELAIS source (ELAISR163442+410759) with a galaxy ID with associated P_{ran} = 1.000, it is very faint and far from the source. There is, however, another ELAIS source (ELAISR163445+410817) present, with detected emission only in the radio, in the error ellipse without an optical ID which is only about 5arcsec from our FIRBACK optical ID.

FN2 - 018: Identification with an optical galaxy (P_{ran} = 0.037, r' magnitude = 19.5, z_{phot} = 0.15, SO/a). However, the 175 μm flux has been assigned in the ELAIS catalogue to a source (ELAISC15_J163334.1+410139) with a different galaxy ID with P_{ran} = 0.998 (r magnitude = 20.5), it is therefore not a possible ID. This also has an IRAS detection at 60 μm. There is another ELAIS source (ELAISR163333+410112) within the FIRBACK error ellipse whose galaxy ID has a P_{ran} = 1.000

FN2 - 019: Identification with an optical galaxy (P_{ran} = 0.001, r' magnitude = 18.1, z_{phot} = 0.10, Sb). The ELAIS source (ELAISR163716+404825) to which the 175 μm flux has been assigned does not have an optical ID, it has detected emission at 6.7μm and 20 cm. This also has an IRAS detection at 60 μm.

FN2 - 020: Identification with an optical galaxy (P_{ran} = 0.086, r' magnitude = 20.2, z_{phot} = 0.15). The ELAIS source (ELAISC15_J163242.7+410627) to which the 175 μm flux has been assigned is fainter (r magnitude = 22.1) and further from the source than our ID and therefore has a higher P_{ran} value of 0.990.

FN2 - 021: Identification with an optical galaxy (P_{ran} = 0.001, r' magnitude = 17.6, z_{phot} = 0.10). However, the flux of this FIRBACK source is below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

FN2 - 022: Identification with an optical galaxy (P_{ran} = 0.019, r' magnitude = 19.1). The ELAIS source (ELAISC15_J163708.1+412856) to which the 175 μm flux has been assigned has an optical identification with a bright optical galaxy with r' magnitude = 18.8, it has P_{ran} = 0.140 and is therefore a possible identification for the FIRBACK source. This ELAIS source has z_{spec} = 0.17, is an Sa and detected emission at 15, 6.7 μm and radio emission at 20 cm. There is another ELAIS source (ELAISC7_J163709+412832) within the FIRBACK error elipse, however, its optical ID is stellar and therefore it has P_{ran} = 1.0.

FN2 - 023: Identification with an optical galaxy (P_{ran} = 0.018, r' magnitude = 19.0, z_{phot} = 0.02, Sb/c). However, the flux of this FIRBACK source is below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

FN2 - 024: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

FN2 - 025: Identification with an optical galaxy (P_{ran} = 0.119, r' magnitude = 20.2, z_{phot} = 0.20, Sa).The ELAIS source (ELAISC7_J163628+404757) to which the 175 μm flux has been assigned has optical ID classed as a galaxy, however, it looks stellar and has a P_{ran} of 0.974 so is not a possible ID. There is another ELAIS source (ELAISC7_J163633+404749) within the FIRBACK error elipse, however, its optical ID is stellar and therefore it has P_{ran} = 1.0. This also has IRAS detections at 60 and 100 μm.

FN2 - 026: This source does not have a confident optical ID. It is associated with a galaxy with r' magnitude 22.4 with a P_{ran} = 0.906. There is an ELAIS source (ELAISC15_J163615.7+404759) within the FIRBACK error ellipse, however, it has a stellar optical ID and therefore has P_{ran} = 1.0.

CFN2 - 027: This source does not have a confident optical ID. It is associated with a galaxy with r' magnitude 20.7 with a P_{ran} = 0.305. This does not agree with the ELAIS source (ELAISR163703+412425) optical ID which is a galaxy with P_{ran} = 0.999.

CFN2 - 028: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 029: Identification with an optical galaxy (P_{ran} = 0.070, r' magnitude = 19.6, z_{phot} = 0.15). The 175 μm flux has been assigned to a fainter galaxy (ELAISR163418+410729, r' magnitude = 22.2) which is further from the FIRBACK position than our ID, it therefore has a high P_{ran} of 1.0. This also has an IRAS detection at 60 μm. There are three other ELAIS sources (ELAISR163419+410641, ELAISC15_J163421.4+410622, ELAISR163422+410648) within the FIRBACK error ellipse, however, none of them have optical IDs.

CFN2 - 030: Identification with an optical galaxy (P_{ran} = 0.086, r' magnitude = 20.3, z_{phot} = 0.15, S0/a). However, the flux of this FIRBACK source is below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue. However, it has IRAS detections at 60 and 100 μm.

CFN2 - 031: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 032: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 033: This source does not have a confident optical ID. It is associated with a galaxy with r' magnitude 21.8 with a P_{ran} = 0.934.

CFN2 - 034: Identification with an optical galaxy (P_{ran} = 0.030, r' magnitude = 19.4, z_{phot} = 0.55, S0/a). However, the flux of this FIRBACK source is below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 035: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 036: Identification with an optical galaxy (P_{ran} = 0.119, r' magnitude = 19.4, z_{phot} = 0.10, Sa). However, the flux of this FIRBACK source is below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue. However, it has IRAS detections at 60 and 100 μm.

CFN2 - 037: This source does not have a confident optical ID. It is associated with a galaxy with r' magnitude 22.1 with a P_{ran} = 0.812. When associations are allowed with all objects this source is identified as a star with P_{ran} = 0.183. However, this still does not agree with the stellar optical ID of the ELAIS source (ELAISR163812+405452) assigned the 175 μm flux.

CFN2 - 038: Identification with an optical galaxy (P_{ran} = 0.023, r' magnitude = 19.3, z_{spec} = 0.14, Sb/c). This is also the optical ID for an ELAIS source (ELAISC15_J163431.5+412246) with detected emission at 15 μm and has an IRAS detection at 100 μm.

CFN2 - 039: Identification with an optical galaxy (P_{ran} = 0.001, r' magnitude = 17.2, z_{spec} = 0.07, Sb). This is also the optical ID for an ELAIS source (ELAISC15_J163613.6+404230) with detected emission at 15, 6.7, 90 μm and radio emission at 20 cm. This also has an IRAS detection at 60 μm.

CFN2 - 040: This source does not have a confident optical ID. It is best associated with a galaxy with r' magnitude 20.5 and P_{ran} = 0.192. The ELAIS source (ELAISC15_J163641.1+413131) to which the 175 μm flux has been assigned has a galaxy optical identification with P_{ran} = 0.562. There is another ELAIS source (ELAISC15_J163648.1+413134) within the FIRBACK error ellipse with a galaxy optical ID with P_{ran} = 0.393.

CFN2 - 041: This source does not have a confident optical ID, however, its best association agrees with the optical identification of the ELAIS source (ELAISC15_J163433.6+405953) to which the 175 μm flux has been assigned. This galaxy has r' magnitude = 19.2 and P_{ran} = 0.311, the ELAIS source has detected emission at 15 μm and 20 cm. There are two other ELAIS sources (ELAISR163427+405936T, ELAISC15_J163430.1+410055) within the FIRBACK error ellipse, one has a galaxy optical ID with P_{ran} = 0.998 and the other does not have an optical identification.

CFN2 - 042: This source does not have a confident galaxy optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 043: This source does not have a confident optical identification when associations are only allowed with galaxies (P_{ran} = 0.437). However, when all objects are possible associations it is matched with a star (P_{ran} = 0.056) which is also the ELAIS source (ELAISC7_J163546+404929) ID to which the 175 μm flux has been assigned. This ELAIS source has detected emission at 6.7 μm.

CFN2 - 044: Identification with an optical galaxy (P_{ran} = 0.053, r' magnitude = 20.0, z_{phot} = 0.05, S0/a). The ELAIS source (ELAISC15_J163730.4+404542) to which the 175 μm flux has been assigned has a galaxy ID with a P_{ran} value of 0.995. This also has an IRAS detection at 60 μm. There is another ELAIS source (ELAISR163728+404533) within the FIRBACK error ellipse, however, this source does not have an optical ID.

CFN2 - 045: This source does not have a confident galaxy optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 046: This source does not have a confident optical association. The flux of this FIRBACK source is also below the ELAIS catalogue 5σ limit of 223 mJy and therefore as it has no associations at other ELAIS wavelengths it is omitted from the catalogue.

CFN2 - 047: Identification with an optical galaxy (P_{ran} = 0.070, r' magnitude = 19.0). The 175 μm flux has been assigned to an ELAIS source (ELAISC15_J163449.5+412048) with a galaxy optical ID with P_{ran} = 0.125 which is therefore a possible association. This source has z_{spec} = 0.25, is an S0/a and detections at 15, 6.7 μm and 20 cm. There is another ELAIS source (ELAISC15_J163451.9+411944) within the FIRBACK error ellipse with a galaxy optical ID with P_{ran} = 0.911.

CFN2 - 048: This source does not have a confident optical identification, is is best associated with a galaxy with P_{ran} = 0.920. The ELAIS source (ELAISC15_J163739.2+405643) to which the 175 μm flux has been assigned does not have an optical identification.

CFN2 - 049: Identification with an optical galaxy (P_{ran} = 0.092, r' magnitude = 20.3, z_{phot} = 0.18). An ELAIS source (ELAISC15_J163741.3+411913) ID has an optical ID with the same galaxy, however, they are different entries in the INT catalogue caused by the splitting up of bright objects by the image analyser. The ELAIS source has detected emission at 15 and 90 μm and is possible an Sa. This also has an IRAS detection at 60 μm. There is another ELAIS source (ELAISC15_J163738.9+411840) within the FIRBACK error ellipse with a galaxy optical ID with P_{ran} = 0.911.

CFN2 - 050: This source falls between two WFS chips.

CFN2 - 051: This source does not have a confident optical identification, it is best associated with a galaxy with P_{ran} = 0.866. The ELAIS source (ELAISC15_J163616.2+411426) to which the 175 μm flux has been assigned has a stellar optical ID and therefore is not a possible ID for the FIRBACK source.

CFN2 - 052: This source does not have a confident optical identification, it is best associated with a galaxy with P_{ran} = 0.466. The ELAIS source (ELAISC15_J163403.0+410350) to which the 175 μm flux has been assigned has a galaxy optical ID with P_{ran} = 0.962. There are two other ELAIS sources (ELAISC7_J163407+410219, ELAISR163410+410326) within the FIRBACK error ellipse one with a stellar ID and one with a galaxy ID, however, both have P_{ran} = 1.0.

CFN2 - 053: This source does not have a confident optical identification, it is best associated with a galaxy with P_{ran} = 0.441. The ELAIS source (ELAISR163621+412244) to which the 175 μm flux has been assigned has a galaxy optical ID with P_{ran} = 0.989.

CFN2 - 054: This source does not have a confident optical identification. When associated with all optical objects it is confidently associated with a star with P_{ran} = 0.215, the ELAIS source (ELAISC7_J163658+411417) assigned the 175 μm flux also has a stellar ID with P_{ran} = 0.350 which is below the limit of 0.4 and therefore a possible association.

Astronomy & Astrophysics manuscript no. paperN105 April 18, 2007
(DOI: will be inserted by hand later)

FIRBACK IV. Towards the nature of the 170 μm source population

Dennefeld M.[1⋆], Lagache G.[2], Mei S.[2,3], Ciliegi P.[4], Dole H.[2], Mann R.G.[5], Taylor E.L.[5], Vaccari M.[6]

[1] Institut d'Astrophysique de Paris, 98bis Boulevard Arago, F-75014 Paris
[2] Institut d'Astrophysique Spatiale, Bât. 121, Université Paris XI, F-91405 Orsay
[3] Johns Hopkins University, 3400 N. Charles Street, 21218, Baltimore, MD, USA
[4] INAF-Osservatorio Astronomico di Bologna, via Ranzani 1, I-40127 Bologna
[5] Institute for Astronomy, University of Edinburg, Royal Observatory , Blackford Hill, Edinburgh, EH9 3HJ United Kingdom
[6] Astrophysics Group, Blackett Laboratory, Imperial College, Prince Consort Road, SW7 2AZ London

Received 13 August 2004; Accepted 14 march 2005

Abstract. We present a detailed study of the brighter ($> 4\sigma$ detections) sources in the 170μm FIRBACK northern N1 ISO survey, with the help of complementary data in the optical, radio, and mid-IR domain. For 82% of them, an optical galaxy counterpart is identified, either as the unique source of the IR emission, or as part of a multiple identification. With less than 15% of AGNs, these sources are essentially local, moderate starbursters with a dominating cold dust component. They are therefore very similar to the galaxies in the IRAS Very Faint Survey or the ISO 170μm Serendipity Survey, and represent a population of cold galaxies rather neglected up to now. Their colours do not match those of the far-IR Cosmic IR Background (CIB), to which they contribute less than 5%. The bulk of the sources contributing to the CIB is thus to be searched for in more distant galaxies, possibly counterparts of the fainter FIRBACK sources still under study. These bright, local, galaxies however play an important role in the evolution of IR galaxies: they dominate the number counts at high 170 μm fluxes, and represent half of the contribution at 250 mJy. Although not particularly massive (typically M*), they form more stars than a typical spiral galaxy and many are bulge dominated, that could represent the remnant of a former merger. The fainter part of this population may represent the missing link with the higher-z sources found in sub-mm observations.

Key words.

1. Introduction

The spectral shape of the Cosmic Infrared Background (CIB) detected by the COBE satellite (*FIRAS* experiment between ~150 μm and ~1 mm, and *DIRBE* experiment at 100, 140 and 240 μm) (Puget et al. (1996), see Hauser & Dwek (2002), for a review) indicates a peak at $\sim$ 150 μm with his energy density at least comparable to the optical/UV background. This peak arises from absorbed optical/UV radiation from star formation and AGN activity in obscured galaxies, that is re-radiated in the far-infrared. This obscured population of galaxies could host approximately half of the massive star formation activity over the history of the Universe (e.g Flores et al. 1999).

Together with the Lockman Hole survey (Kawara et al. (1998)), the FIRBACK survey (Dole et al. 2001) was the most reliable and deepest (σ(170 μm) $\sim$ 45 mJy) infrared census at wavelengths between 20 μm and 850 μm, and the only far-infrared survey with a better sensitivity than IRAS, until *SPITZER* observations become available. It is close in wavelength to the peak of the CIB seen by COBE, and because of this wavelength proximity, characterizing the FIRBACK sources is an essential step towards understanding the nature of the sources that generate the CIB.

The FIRBACK survey used about 150 hours of ISO observing time, corresponding to one of the largest ISO programs (Kessler et al. 2000). It covers about 4 square degrees in three high galactic latitude fields, called FIRBACK South-Marano (FSM), FIRBACK North 1 (FN1) and FIRBACK North 2 (FN2). The northern fields observed in FIRBACK are a subset of the larger area covered by the ELAIS survey at shorter wavelength (Oliver et al. 2000). The precise location of the fields, the details of the reduction process, an assessment of the reliability of the detected individual sources, as well as the positions of the close to 200 sources detected with flux S>135 mJy (the 3σ limit) are given in Dole et al. 2001. Our knowledge of these sources is still quite limited. Patris et al. (2003) have published spectra of the 21 brightest FIRBACK sources in the southern FSM field. These bright sources are mostly nearby (z<0.2) dusty star forming galaxies. They exhibit star formation rates of a few tens of solar masses per year with typical IR luminosities of about 10^{11} $L_{\odot}$. The fraction of Active Galactic Nuclei (AGN) is low, around 15% at most. In the north, two optical sources (I = 24 & 22) were observed in FN1 with ESI on Keck, revealing higher redshift objects (z = 0.5 and 0.9, respectively (Chapman et al. 2002). These two sources are identified as UltraLuminous IR galaxies (ULIRG's), with merger morphologies and relatively cold dust temperatures.

Send offprint requests to: dennefel@iap.fr

* Visiting Astronomer, Observatoire de Haute-Provence, CNRS, France

Finally, based on a statistical analysis (Sajina et al. (2003)), the FN1 sources appear to show a bimodal redshift distribution, with normal star-forming galaxies at $z \simeq 0$ and a tail of a much more luminous galaxy population at $z \sim 0.4$-0.9 .

In this paper, we present the information available in the FN1 field, for all FIRBACK sources with flux $S > 4\sigma$ (=185 mJy) and give some insights on the fainter flux population. The detailed analysis of all the FIRBACK population is not complete up to now: the identification process is a long process which requires many complementary observations in various wavelength ranges. This paper describes the currently available data in this N1 field, thus allowing the community to conduct complementary follow-up observations if appropriate. Emphasis is put on the brighter (thus presumably closer) sources where the identification is secure. Work is continuing on those sources with multiple or fainter optical counterparts, which probably represent a population further away.

The paper is organised as follows. The complementary observations available to date are presented first (Sect. 2). Then, we describe the identification process and give the results source by source (Sect 3.). In Sect. 4 and 5, we discuss the optical and IR properties of our sample. The IR/radio correlation is discussed in Sect. 6, and in Sect. 7, we analyse the star formation rates. We finally give some global properties for the population of sources with fluxes between 3σ and 4σ in Sect. 8 and discuss all the results in the last section (Sect. 9).

2. Complementary observations

The FN1 field has been covered (sometimes in part only) by radio, millimeter, sub-millimeter, 90, 15 and 6.7 μm, observations, U, g', r', i', Z and K-band imaging and visible spectroscopy. In this section, we describe briefly all these data and their relevance for the present analysis.

2.1. The ELAIS observations

The European Large Area ISO Survey (ELAIS) has surveyed a total of 12 square degrees at 15 μm using the ISOCAM instrument (Cesarsky et al. 1996) and at 90 μm using ISOPHOT (Lemke et al. 1996). Seven square degrees have also been covered at 6.7 μm, but only one square degree at 170 μm (the FN2 field). The ELAIS final band-merged catalog has recently become available (Rowan-Robinson et al. 2004).

The FIRBACK-FN1 field discussed here has been covered at 170 μm over two square degrees in an additionnal observing run conducted during the "supplementary" lifetime of the ISO satellite. For association of FIRBACK sources in the FN1 field with shorter wavelength data, only the ELAIS data at 15 μm have been used, as it is the only wave-

length where the field coverage is nearly complete. In general, we prefer not to use the 90 μm catalogue (Heraudeau et al. 2004) since it is clearly incomplete, although, in some cases, positions of 90 μm sources with high fluxes have been used to make the identification process converge. Since the ELAIS survey is a shallow survey, the sensitivity at 15 μm is limited by the instrumental noise. The catalogue of Vaccari et al. (2005) (http://astro.imperial.ac.uk~vaccari/elais), which is used here, has a completeness limit at the faint end of about 1 mJy at 15 μm, with a 1σ photometric error better than 0.25 mJy. It provides a source density of about 170 sources per square degree on average, down to the 5 σ limit, to ease the FIRBACK identifications. Its rather low sensitivity biases the 15 μm detections towards the lowest redshift sources, but corresponds well to the brigthest FIRBACK population.

2.2. The radio data

The Very Large Array (VLA) has been used by Ciliegi et al. (1999) in C configuration to carry-out a sensitive 20-cm radio survey of the FN1, N2/ELAIS and N3/ELAIS fields. In the FN1 field, the flux limit varies over the area surveyed: a circular central region with a flat noise distribution (~200 arcmin^2, 5σ=0.135 mJy), surrounded by concentric annular regions, where the noise increases for increasing distance from the centre (the last annulus has a 5σ rms noise of 1.15 mJy). These regions of the sky had also been surveyed previously to shallower flux limits at 20cm, in the NVSS survey in the VLA D configuration (Condon et al. (1998)) and the FIRST survey in the VLA B configuration (Becker et al. (1995)).

For the identification process, we use both the 5σ radio catalogue of Ciliegi et al. (1999) and the FIRST catalogue. When no radio source is found in the FN1 position error circle, we go back to the radio map to compute 5σ upper limits and search for radio sources down to the local 3σ noise value. Within the error circles of the 56 brightest FIRBACK FN1 sources, we found at least one radio source for 30 of them. Eight of those have their radio flux in the $3 < S < 5\sigma$ range, and three have been found in the FIRST catalogue. When the FN1 source position is outside the Ciliegi et al. (1999) coverage, a 5σ upper limit from the FIRST survey is used.

2.3. Millimeter and sub-millimeter data

IRAM 1.3 mm observations have been conducted with MAMBO on some (~10) radio positions that were coincident with FIRBACK positions. Only 2 or 3 sources seem to be detected. The data are currently under analysis and will not be used in this paper.

On the sub-millimeter side, early observations of some FIRBACK sources were published by Scott et al. (2000), followed by observations of a sample of 30 FIRBACK N1 sources by Sajina et al. (2003). This sample as a whole (co-adding all the data for all the observed sources) is detected at the 10.6σ level at 850 μm and at the 9.0σ level at 450 μm ($< S_{850} >= 2.5 \pm 0.2$ mJy and $< S_{450} >= 16.7 \pm 1.9$ mJy). Out of the 30 N1 sources, 7 are detected at the $>3\sigma$ level at 850 μm (3 of them have $3\sigma < S < 4\sigma$) and 5 sources are detected at the $>3\sigma$ level at 450 μm. Only these $S > 3\sigma$ detections will be used here.

2.4. X-ray data

A cross-correlation between the ELAIS-ISO 15 μm survey and the ROSAT all-sky survey has been conducted by Basilakos et al. (2002). Three sources were found, corresponding roughly to the fields of the 170 μm sources FN1-35, 126 and 295. But the position of the X-ray source is in these 3 cases always at more than 140" from the centroid of the PHOT source, and is therefore largely outside our error circle. We can therefore conclude that there is no association of a 170 μm source with an X-ray source, at the sensitivity level of the Rosat all-sky survey.

Another cross- correlation of the 15 μm survey with specific Chandra pointings in the N1 (and N2) field has been performed by Manners et al. (2004). Three matches were found in N1, but none correspondings to a FIRBACK source (except perhaps $FN1-042$, see the discussion of individual sources).

2.5. Other IR data.

We derived IRAS 60 and 100 μm fluxes (or upper limits) for each FN1 source position, by using SCANPI[1], a tool developed for visualizing, plotting and averaging calibrated IRAS survey scans. Each FN1 source was checked by eye. For point sources with signal-to-noise greater than 3, we determine the flux by fitting the point source template. When no clear detection was obtained, a 3σ upper limit was set. The whole procedure is fully described at: http://irsa.ipac.caltech.edu/IRASdocs/scanpi/. None of the identified FIRBACK galaxies is resolved by IRAS. We checked that the fluxes derived by using SCANPI (in and near the FIRBACK fields) were in good agreement with those of the IRAS Faint Source Catalog for sources in common.

We also searched for associations of optical counterparts with 2MASS (Kleinmann et al. (1994)) sources and an association was found in the Extended Source Catalogue (XSC) in most cases. Although several magnitudes and colours are usually available, we have extracted only the total K_s magnitude: we do indeed not

[1] http://irsa.ipac.caltech.edu/applications/Scanpi/

intend to use the near-IR colours to fit the spectral energy distributions (in view of the probable complex stellar populations content), but simply use the K luminosity to estimate the mass of the galaxy. In a few cases when no 2MASS magnitude was available, we used K magnitudes from the work of Sajina et al. (2003).

The N1 field is one of the fields of the SWIRE legacy survey (Lonsdale et a. (2003)) using the SPITZER satellite (Werner et al. (2004)) and is the first one to have been observed. The products (Version 1.0) comprise images and catalogues in the four IRAC bands (3.6, 4.5, 5.8 and 8.0 μm) and three MIPS bands (24, 70 and 160 μm), described by Surace et al. (2004) on the Spitzer Web site (http://ssc.spitzer.caltech.edu/legacy/), and were released on Oct. 27, 2004, after this paper was first submitted. The use of those data for the analysis of FIRBACK sources is thus defered to a subsequent paper. We note however that the available MIPS 160 μm catalogue contains 178 entries down to the cut-off limit of 200 mJy, in an area of 8.5 square degrees. Twenty-five of them only lie within the smaller FN1 area, to be compared to 44 FIRBACK sources down to the same flux limit (at 170 instead of 160 μm). We have however checked on the SWIRE 160 μm images that all the FIRBACK sources were present down to 200 mJy, and almost all of them (only three possibly suspicious cases) down to 140 mJy, so that all the FIRBACK sources discussed in this paper are real and that the difference in number counts is to be ascribed to incompleteness of the preliminary SWIRE 160 μm N1 catalogue, and probably incomplete field coverage. While the sensitivities of both surveys are therefore comparable, the SWIRE data will later allow a larger sample of presumably similar objects to be studied, with the help of the shorter wavelength data for more precise identifications.

2.6. Optical data

Multicolour photometry was obtained over the FN1 field, in U through Z, as part of the INT Wide Field Survey [2] (MacMahon et al. (2001)), with the Wide Field Camera covering about 0.3 deg^2 at once on the 2.5m Isaac Newton Telescope. The photometry is described in Gonzalez-Solares et al. (2005) and the limiting magnitudes range from ~21.9 in Z to ~ 24.9 in g'. The corresponding catalogue was used to search for optical associations with the FN1 sources using the likelihood ratio method, as described most recently in Mann et al. (2002). For each source, the probability P_{ran} is computed that an association with an optical object at the given likelihood ratio has occurred by chance, on the basis of simulations using associations with random positions in the INT data. A low P_{ran} value therefore means a more secure identification. The results of these computations are used together with the 15 μm and 21 cm catalogs to find the best

[2] http://www.ast.cam.ac.uk/ mike/casu/WFCsur/WFCsur.html

identifications (as explained in Sect 3).

DSS2 data were used for initial identification purposes and preparation of the spectroscopic observations, waiting the accessibility of the INT images, and are shown as small charts in Fig. 10.

Optical spectra were taken at the Haute-Provence Observatory (CNRS, France) with the 1.93m telescope during several observing runs between 1999 and 2004. The Carelec long-slit spectrograph was used with a 300 l/mm grating and an EEV CCD detector of 2048x1024 pixels of 13.5 μm each, giving a spectral element of 1.75Å per pixel and a spectral resolution of 6.5Å with the 2″ entrance slit generally used. The total spectral coverage is about 3600Å but the central wavelength was different from one run to another, to allow the Hα emission sometimes to be detected in higher redshift objects. The orientation of the slit was adjusted to register more than one object in a given exposure, when adequately bright galaxy candidates were available in the PHOT error circle.

The reduction followed the usual procedure, with flat-fielding, wavelength calibration, and spectral response determination through observations of several standard stars, the master standard usually being BD+26,2606. The relative response is determined with an accuracy of a few percent, the blue part of the spectrum having the poorest correction due both to the decreasing response of the detector and the average extinction curve available. Absolute fluxes are available only part of the time, due to changing weather conditions: when they are, their accuracy is estimated to be better than 15%, not taking into account light losses at the entrance slit. As these fluxes are only indicative, no correction for those losses has been attempted.

3. Identifications: method and results

As the FIRBACK position error circle (Dole et al. 2001) is rather large (100 ″ in diameter for a 93% localising probability), the identification process is complex. For the bright PHOT sources (most of those considered in this paper), we selected the bright galaxy (or galaxies) within the PHOT error circle and used, as the major criterion, the distance between the optical galaxy candidate and the centroid of the PHOT detection. If several candidates were possible, the spectral characteristics were then used to identify the good ones: the FIRBACK counterparts are expected, by analogy with IRAS sources, to be primarily starforming galaxies (or AGNs), albeit possibly colder as they are selected at longer wavelengths, and their spectral features should thus be rather similar to the typical objects found in the IRAS survey (e.g. Veilleux et al. (1995) and references therein). The complementary observations with better positional accuracy were used conjointly to find the correct optical association(s) to the far-IR source. We

primarily used the ISOCAM 15 μm and/or the radio 21 cm data. The ratio of sensitivity between ISOCAM at 15 μm (which is instrumental noise limited) and ISOPHOT at 170 μm (which is confusion noise limited) is such that it is unlikely that an extragalactic source will be detected at 15 μm without contributing significantly to the FIRBACK 170 μm flux (unless it has a very unusual SED). On the other hand, the use of the radio data assumes that the detected sources have a similar radio to far-IR ratio as sources previously detected by IRAS (e.g. Condon & Broderick (1991)). The use of this criterion could prevent the detection of a different type of source in this sample (should they exist), but in practice, at least for the bright optical galaxies observed here, there was in general no ambiguity: the bright galaxy detected closest to the PHOT source indeed had spectral characteristics similar to those of IRAS galaxies; it moreover very often had an associated 15 μm detection so that the use of the radio data was not essential. It will however become important when going to fainter PHOT sources, where more than one, faint, optical counterpart can be located within the error circle.

FIRBACK sources were then distributed in three groups:

1. Identified sources: sources associated with one bright optical galaxy, usually also detected in radio and/or ISOCAM. Fainter optical sources, like FN1-040 (which lies at z=0.45), are also put in this category, when the radio/ISOCAM identification was also unambiguous.
2. Multiple sources or uncomplete identification. Into this category are put FIRBACK sources with more than one optical/15 μm/radio counterpart and FIRBACK sources with no secure identification.
3. Unidentified sources. Either no radio or 15 μm source detected to help identification, or the optical galaxies are too faint to be detected spectroscopically with the available instrument. We cannot reach a conclusion in such cases.

Out of the 56 ISOPHOT sources in the 4 σ sample, 28 are fully identified, 17 have multiple or incomplete identifications and 11 sources have no identified counterpart. We comment on a few sources in the $3 < S < 4\sigma$ range, which were also analysed. Fig. 10 presents the DSS images with the FIRBACK error circle and the radio and ISOCAM counterparts when present.

3.1. Identified sources

This section gives details concerning the identified sources. We emphasize the complementary data (near-IR or radio) when they are important to secure the identification. For all sources, fluxes in the various bands (IR and radio) and far-infrared colors are given in Table 1. Optical properties are given in Table 2 and spectral characteristics are given in Table 4.

*FN*1 – 000: Identification with a bright optical galaxy, also detected at 15 and 90 μm. This galaxy is a merger, with clear tidal tails. The two other bright objects inside the PHOT error circle are stars.

*FN*1 – 001: Bright optical galaxy.

*FN*1 – 002: Bright optical galaxy, also detected at 90 μm.

*FN*1 – 003: Bright optical galaxy with 15 μm emission. This identification has a higher P_{ran} probability to occur by chance, because the galaxy is located $\approx$30 $''$ from the ISOPHOT centroid but there are no other obvious identifications on the image. This source is detected by IRAS at 100 μm only (upper limit at 60 μm, although this band is more sensitive) and should therefore be a rather cold galaxy. The spectrum shows a low equivalent width of Hα, suggesting a weak starburst in an older galaxy.

*FN*1 – 004: Bright optical galaxy with a point-like nucleus. This source is also detected by SCUBA at 450 μm (32.5$\pm$7.1 mJy). Its spectrum reveals a Seyfert 2 type (strong, narrow [NII] lines). A fainter galaxy at the southern edge of the error circle has a similar redshift and also Hα emission, but is not detected in radio nor at 15 μm: it is unlikely to contribute much to the far-IR flux.

*FN*1 – 005: Faint optical galaxy with 15 μm detection. The identification has a higher P_{ran} value, because the galaxy is located $\approx$30 $''$ from the ISOPHOT centroid, but the ISOCAM identification is unambiguous. This object has been observed in the optical with the Palomar200/DoubleSpec instrument (Chapman, private communication).

*FN*1 – 006: bright optical galaxy detected at 15 μm and 90 μm. Because of its location $\approx$30 $''$ away from the PHOT centroid, this identification would have a higher probability of occuring by chance, but the spectrum is typical of a reddened starburst. The bright object just east of the galaxy is apparently a star. A second 15 μm source, much fainter (a factor of about 6) than the other one, is also located within the error circle, but does not show any obvious optical counterpart. In view of its faintness, and because it does not correspond to the 90 μm source, it is unlikely to contribute much to the 170 μm flux.

*FN*1 – 007: Bright optical galaxy with 15 μm, and 1.4GHz emission. This source is detected by SCUBA both at 850 μm and at 450μm (23.4$\pm$8.1 mJy for the latter). It has a very cold IRAS colour, but does not lie in the few, very faint, well identified cirrus filaments in the N1 field, so is unlikely to be contaminated by background.

*FN*1 – 009: Identification with a bright optical galaxy with 15 μm and 1.4GHz emission. The spectrum, although dominated by stellar features, shows clearly the presence of an active nucleus (strong [OI], large [NII]/Hα and [SII]/Hα ratios), classified as Sey2. The bright object at the southern edge of the PHOT error circle, with a 15 μm detection, is a spectroscopically confirmed cold star.

*FN*1 – 011: bright optical galaxy with a 15 μm detection. The other bright object to the SW is a star.

$FN1-012$: Bright optical galaxy detected at 15 μm and 90 μm. Although the galaxy is $\approx$30 $''$ from the ISOPHOT centroid, the identification is secure, with a typical reddened starburst spectrum.

$FN1-014$: This source does not have 1.4GHz data (it lies outside the VLA surveyed area) but has a clear 15 μm counterpart and is also detected at 90 μm. The optical objects were however too faint to be observed spectroscopically with our equipment. A photometric redshift has been provided by T. Babbedge (private communication).

$FN1-015$: This source is identified with a bright optical galaxy with 1.4GHz emission, and SCUBA detection. It has one of the highest redshifts measured in our sample, so that Hα falls outside our prime spectral range.

$FN1-016$: Identification with a bright optical galaxy detected at 15 μm, and 1.4GHz. This source is detected by IRAS at 60 and 100 μm, and by SCUBA both at 450 and 850 μm. Although its position is rather offset with respect to the PHOT centroid, the identification seems to be fairly secure, in view of the detection in all these wavebands, and its typical reddened starburst spectrum.

$FN1-018$: bright optical galaxy with 15 μm detection but no radio emission. Its spectrum, with a strong [NII]/Hα ratio, indicates the presence of an AGN.

$FN1-020$: Identification with a faint optical galaxy also detected at 15 μm. This source is detected by IRAS at 100 μm only, and is thus presumably cold. We could not secure an optical spectrum, but a photometric redshift has been provided by T. Babbedge (private communication). The other bright object in the error circle seems to be a star.

$FN1-021$: bright optical galaxy also detected at 15 μm. A second galaxy, at the eastern edge (but outside, at 69") of the PHOT error circle, also detected at 15 μm, with a starburst-type spectrum and a velocity close to the bright galaxy one, might also contribute to the detected far-IR flux.

$FN1-023$: bright optical galaxy with a 15 μm detection, but not detected by IRAS. The bright object NE of the galaxy is starlike.

$FN1-024$: Bright optical galaxy with 15 μm and 1.4GHz emission and a starburst spectrum. Although there are other, fainter galaxies in the field (but not detected in radio nor in mid-IR), this source is considered as the main contributor to the far-IR emission.

$FN1-026$: Identified with an optical galaxy having also 15 μm and 90 μm detections, but no radio detection. The other two bright objects in the error circle are presumably stars.

$FN1-031$: Identified with an optical galaxy with 90 μm and 1.4GHz detections. This source is also detected by SCUBA at 850 μm.

$FN1-033$: Identified with a faint optical galaxy with 15 μm emission, but no radio detection.

$FN1-035$: Optical galaxy with also a 15 μm detection but no radio detection (2MASS

detection, galaxy classified as IrS).

*FN*1 – 038: The faint optical galaxy inside the error circle displays emission lines in its spectrum and is associated with a 15 μm source. A radio source exists, with no obvious optical counterpart, but lies outside the error circle and is therefore not associated with the PHOT source.

*FN*1 – 039: Identified with a faint optical galaxy detected at 1.4GHz. This source is also detected by SCUBA at 450 μm, and has been spectroscopically observed with the Palomar200/DoubleSpec instrument (Chapman, private communication).

*FN*1 – 040: Identified with a faint optical galaxy detected at 1.4GHz and with SCUBA at 850 μm. The bright object in the center of the error circle is a star. This is one of the 2 higher redshift FIRBACK sources detected by Chapman et al. (2002), who note it is an interacting pair.

*FN*1 – 041: Identified with a bright galaxy in the center, detected in radio and in mid-IR, with a typical emission-lines spectrum.

*FN*1 – 043: Identified with an optical galaxy, also detected at 15 μm. A high [NII]/Hα ratio indicates an active nucleus, but the S/N of the spectrum is insufficient to distinguish between a liner or a reddened Sey2.

3.2. *Multiple or incomplete identifications*

We detail in this section the FIRBACK sources with more than one possible identification and/or a non-secure identification.

*FN*1 – 008: Two sources with 1.4GHz emission within the error circle. The brightest optical counterpart is at z=0.26 and has a 350 μm CSO detection. It shows a reddened spectrum with absorption features due to hot stars, but Hα is unfortunately out of our spectral range. The faintest galaxy (the second radio source) has a SCUBA measurement and a K-band magnitude of 14.2 (Sajina et al. 2003)

*FN*1–010: This source has two 15 μm counterparts associated with bright optical galaxies. It is also detected by IRAS at 60 μm. The bright galaxy in the error circle has a reddened continuum with strong stellar absorption features, and Hα and [NII] emission of small equivalent width. The [NII]/Hα ratio clearly indicates an AGN, but the non-detection of the emission features in the blue spectral range does not allow us to distinguish between its being a Sey2 or a Liner. The fainter galaxy within the error circle is detected in radio at 1.4GHz and also at 15 and 90 μm (while the bright galaxy is not), but was to faint to be observed spectroscopically here. The identification of the PHOT source is therefore still uncertain, and the far- IR flux may in fact be due to contributions of both galaxies.

*FN*1 – 017: This source has a 15 μm counterpart associated with a faint galaxy located

at the edge of the error circle. Many other faint galaxies are visible in the error circle, but they were all too faint to be observed here.

$FN1-019$: Two possible counterparts: one bright optical galaxy with 15 and 90 μm emission, whose spectrum indicates the possible presence of a Liner (the emission lines have also a small equivalent width). And one fainter galaxy with 1.4 GHz emission (3-5σ detection), but no available spectrum.

$FN1-025$: Two bright galaxies in the error circle display emission lines in their spectrum, but are not related to each other. One has a 15 μm counterpart. A fainter galaxy in the field, this one with 1.4 GHz emission (3-5σ detection), could not be observed spectroscopically. The situation is complex, as all three galaxies may contribute to the far-IR flux.

$FN1-029$: This source could be identified with an optical galaxy close to the PHOT centroid and both 15 μm and 1.4 GHz detections. Another faint, optical galaxy, also with 1.4 GHz emission, lies however at the border (51") of the error circle (not observed spectroscopically) and could contribute also to the far-IR emission.

$FN1-032$: A bright galaxy with 15 μm counterpart, close to the PHOT centroid, displays weak emission lines superposed on a strong, reddened, stellar continuum. The [NII]/Hα ratio, close to 1, indicates the presence of an active nucleus. The 1.4 GHz emission is however associated with another, faint optical galaxy, also close to the center of the error circle, but which was too faint to be observed spectroscopically here.

$FN1-034$: The bright galaxy close to the PHOT centroid displays emission lines in its spectrum, has 15μm emission but no radio counterpart. It is also detected by IRAS at both 60 and 100 μm. The radio 1.4GHz emission in this field corresponds to a very faint optical galaxy (K=19.3), with a SCUBA detection at 450 μm (but not at 850 μm), which could not be observed spectroscopically.

$FN1-037$: faint optical galaxy, with also a 15 μm detection. Could not be detected spectroscopically. Another 15 μm source, inside the error circle, is associated with an even fainter galaxy.

$FN1-042$: The bright optical galaxy close to the center (detected in mid-IR but not in radio) shows only weak emission lines in its spectrum and is unlikely to be the real counterpart. Three radio sources are detected in the vicinity, two inside the error circle, associated with faint galaxies which could not be observed in the optical, and correspond probably to the emitters of the far-IR flux. The third, south-eastern, radio-source is slightly outside the PHOT error-circle and corresponds to one of the X-rays sources found by Manners et al. (2004): they do classify this object as starburst-galaxy, and it could contribute also to the far-IR flux.

$FN1-044$: A comparatively bright optical galaxy, detected in radio at 1.4 GHz, and also at 15 μm and 90 μm with ISO and 60 μm with IRAS, is undoubtedly a counterpart to the 170 μm source. Its spectrum indicates the presence of an AGN (liner or Sey2).

One other radio source is however also present within the error circle, and is associated with a very faint optical galaxy, which could contribute also to the far-IR emission.

*FN*1 – 045: The bright galaxy, with radio emission (and a 2σ SCUBA detection) is the probable identification. However, there are other, fainter galaxies with unknown properties in the field.

*FN*1 – 046: Two bright galaxies are close to the border of the FIRBACK error circle. The one inside, with no radio emission but 15 μm emission, is a standard emission line galaxy which would be considered as the far-IR counterpart if it would be alone. The second galaxy, just outside the error circle, does however not show emission lines. There is moreover another 15 μm source lying in the error circle. This source is associated with a very faint optical galaxy. The identification is therefore not complete here.

*FN*1 – 048: A bright galaxy, with emission lines, is situated close to the center of the error circle. Two radio sources, however, are also present (closer to the boundary), one superposed on a bright star (spectroscopically confirmed as late type), the other on a very faint optical galaxy. This latter source is also detected by SCUBA both at 450 and 850μm, and is likely also to be a contributor to the far-IR source.

*FN*1 – 049: A bright optical galaxy detected at 15 μm and at 1.4 GHz (3-5σ detection) is certainly a contributor to the far-IR source. In the error cicle lies another 15 μm source associated with a very faint optical object.

*FN*1 – 050: Two 15 μm sources are detected at the limit of the error circle. They are both associated with very faint optical galaxies.

*FN*1 – 053: One 15 μm source is detected at the border of the error circle (at 49.7 arcsec). It is associated with a very faint optical galaxy which could not be observed spectroscopically.

ID	ISO 15 μm	IRAS 60 μm	IRAS 100 μm	ISO 170 μm	SCUBA 850 μm	VLA 1.4 GHz	100/60	170/100
FN1-000	27.42±4.17	620± 28	820± 63	837.7± 89.7		4.10^F ± 0.15	1.32	1.02
FN1-001	17.45±2.68	260±31	410 ±99	597±72.5	6.1 ±1.6	0.74 ±0.23	1.58	1.46
FN1-002	11.67±1.87	160±25	340±93	544.5±68.8	4.4±1.1	0.64±0.04	2	1.6
FN1-003	8.94±1.37	< 100	220 ± 60	408.3 ± 59		<0.75	>2.2	1.86
FN1-004	11.38±1.85	100±22	300±73	390.8±57.7	3.6±1.4	0.88±0.10	3.0	1.3
FN1-005	4.15±0.65	110± 34	300± 78	373.8± 56.6		<1.00^F	2.74	1.25
FN1-006	11.6±1.78	100± 29	300±82	347.6± 54.7		<1.00^F	3	1.16
FN1-007	5.74±0.9	70± 25	520± 90	337.6± 54	4.4± 1.6	1.04± 0.10	7.4	0.65
FN1-009	5.71±0.88	<90	< 400	313.1 ± 52.3	3.5±1.5	1.15 ± 0.07		>0.78
FN1-011	2.78±0.47	<78	<255	304.5± 51.6		<1.00^F		>1.19
FN1-012	3.93±0.61	90 ± 34	<400	302.1± 51.4	1.5±1.6	0.31± 0.07	<4.44	>0.76
FN1-014	2.84±0.48	150±19	380±57	295.2±50.9		<1.00^F	2.53	0.78
FN1-015	2.9±0.48	<78	<300	294.4±50.9	1.4±1.6	0.52±0.07	>3.77	>0.98
FN1-016	7.87±1.20	160± 17	360± 92	289.3± 50.5	1.5±1.2	1.55± 0.07	2.25	0.8
FN1-018	2.77±0.46	<105	<210	287.8± 50.4		<1.00^F		>1.37
FN1-020	2.05±0.37	< 108.	230.±76	283.40 ± 50.1		< 0.5	>2.62	1.23
FN1-021	2.98±0.35	130±26	<360	271.3± 49.2		< 0.75	< 2.77	>0.75
FN1-023	3.44±0.55	< 69	<294	269.9 ± 49.2		<1.00		> 0.92
FN1-024	5.02±0.78	<84	<147	266.2±48.9	2.3±1.3	0.75±0.04		>1.81
FN1-026	9.16±0.23	< 66	<243	240.6 ± 47.1		<1.00		>0.99
FN1-031	3.74±0.59	<69	<237	225.2 ± 46	1.9±1.1	0.43 ± 0.04		>0.95
FN1-033	3.62±0.43	< 90.	< 330	224.1 ±45.9		< 0.5		>0.68
FN1-035	4.22 ±0.67	< 63.	< 192	218.2± 45.5		<0.75		>1.14
FN1-038	2.17 ±0.38			206.9± 44.7		<1		
FN1-039		< 57	<246	205.2± 44.5	-0.1±2.3	0.58± 0.05		>0.83
FN1-040		< 75.	< 153	204.7 ± 44.5	5.4± 1.1	0.33 ± 0.03		>1.34
FN1-041	4.50±0.72	<69	<243	204.2±44.5	-0.1±2.5	0.76±0.04		>0.84
FN1-043	2.47±0.4	< 78	<225	200.6 ± 44.3		<0.75		>0.89

^{F} : FIRST radio survey

Table 1. Infrared and radio data (all in mJy) for the FIRBACK identified sources, together with infrared colors.

3.3. Unidentified sources

For all these sources (except 4), there is no radio or ISOCAM detection inside the PHOT error circle, and no obvious galaxy counterpart. For sources $FN1-022$, $FN1-027$, $FN1-051$, $FN1-054$, $FN1-055$, we have no further information to give. The 6 last sources are:

$FN1-013$: There is a mid-IR galaxy (F_{15}=1.36±0.3 mJy) associated with a very faint optical galaxy. We have no further information on this object.

$FN1-028$: Two bright optical galaxies, one at the edge (SE) of the error circle and one well inside (NW) have been observed spectroscopically. They have recession velocities of 19921 and 19816 km/sec respectively, but do not display emission lines . There is no radio nor 15 μm detection to help identification, and the two optical galaxies are unlikely the sources of the far-IR emission.

$FN1-030$: The only bright optical source in the error circle is a spectroscopically confirmed star.

$FN1-036$: One ISOCAM position corresponds to a very bright star. Other bright objects seen in the error circle could also be stars (like the other 15 μm faint source).

$FN1-047$: Only stars are seen inside the error circle, the brightest one being also an ISOCAM 15 μm source.

$FN1-052$: This source has one radio counterpart (with radio detection above 3 sigma) but no clear optical identification.

3.4. Some sources fainter than the 4σ limit

For the fainter ISOPHOT flux galaxies (47 sources, $3\sigma < S < 4\sigma$)), radio and mid-infrared counterparts may be expected to become rarer. However, this is only marginally the case. The ISOPHOT sources without any radio and mid-IR counterparts represent 23% of the $S > 4\sigma$ sample and 34% of the $3\sigma < S < 4\sigma$ sample. Among the 47 ISOPHOT sources with fluxes between 3 and 4σ, 12 have a 15 μm counterpart, 12 have a radio counterpart and 7 have both. This tends to show that the catalog is reliable down to the 3σ level. Two examples were observed optically:

$FN1-057$: A bright galaxy inside the error circle is also detected at 15 μm. It displays emission lines and is the likely identification. The source is also strong at both 60 and 100 μm (IRAS detection).

$FN1-101$: Bright galaxy with emission lines, close to the center of the error circle, also detected at 15 and 90 μm. It also coincides with the only radio source in this field, and is therefore the obvious identification of this much fainter 170 μm source.

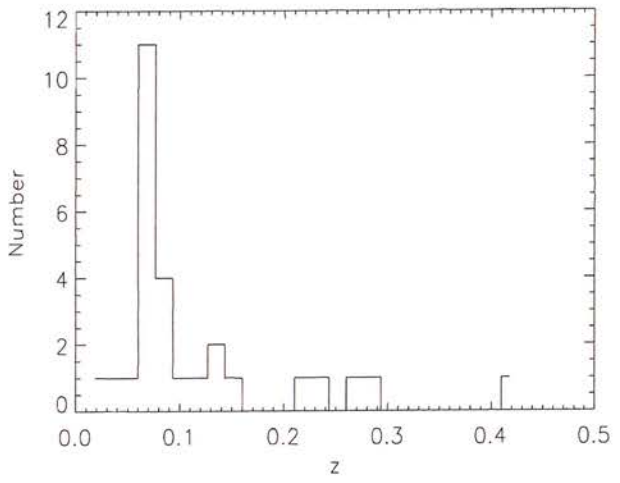

Fig. 1. Redshift distribution of the 28 FIRBACK fully identified sources.

In all the following (except when specifically mentioned), we discusse only sources with secure identifications (28 sources).

4. Optical properties of identified sources

4.1. Optical colors and morphologies from the INT-WFC survey

FIRBACK sources have been observed as part of the INT Wide Field Survey in the U, g', r', i' and Z filters. INT photometric bands are similar to the SDSS bands (Sloan Digital Sky Survey (Fukugita et al. 1996), see the WFS section for further details (Sect. 2.6).

The morphology of all the objects in our sample was assessed visually on the i' images, using the common Hubble sequence as in Postman et al. (2005). With the Advanced Camera for Surveys (ACS) on the Hubble Space Telescope, a visual classification of that kind has a typical random error of 25%, which can be reduced to 6% when only two broad categories are used (early-type, and spirals/irregulars). Although we used ground-based images here, the objects are all relatively close so that the uncertainty in classification is of the same order of magnitude.

Optical colors were derived for identified sources and they are given in Table 2. The source $FN1-040$ is too faint for precise photometry and $FN1-000$ does not have a g' magnitude. They are both excluded from the following analysis. The $(g'-r')$ and the $(U-r')$ color histograms are shown in Fig. 2. Strateva et al. (2001) have found, from the analysis of about 150000 galaxies in the SDSS, that the distribution of galaxies in $(U-r')$ is strongly bimodal with an optimal color separator of $(U-r')=2.22$ for low redshift ($z<0.4$) objects: late type galaxies lie mostly on the blue side, with $(U-r')<2.22$. Although selected from the far-IR, most of our galaxies lie also on the blue side: this is coherent with their classification as late-type for most of them, but suggests that

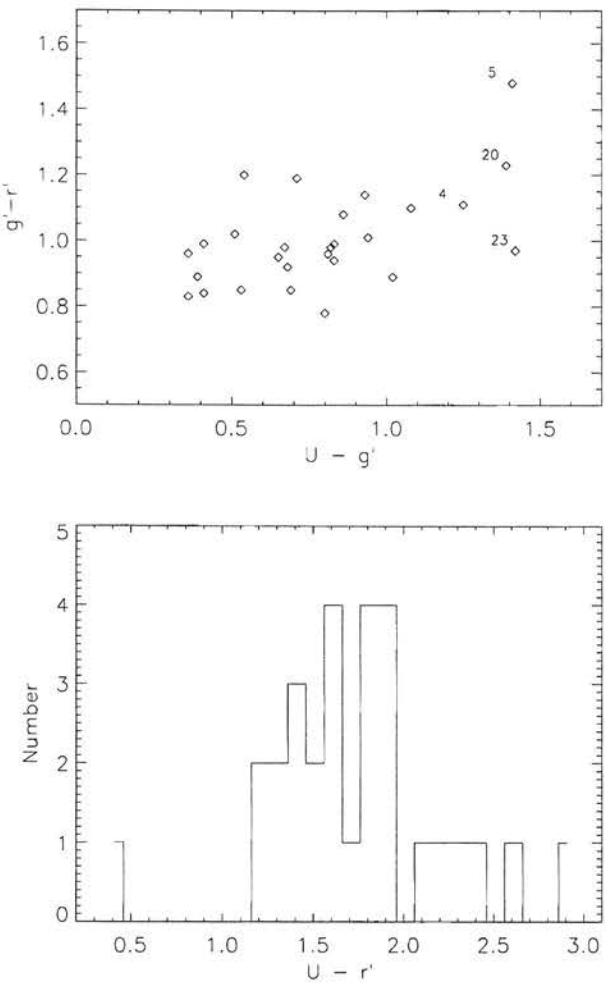

Fig. 2. Optical colors of the identified sources. At the top, we show the (g' - r')-(U - r') colour-colour diagram, while the bottom gives the (U - r') histogram.

they are not heavily obscured, at least outside the central regions. The reddest sources ($(U - r') > 2.2$; FN1-04, FN1-05, FN1-20, FN1-23) are also spiral galaxies, but this time probably affected by heavier extinction.

These measurements have not been corrected for galactic extinction, but this correction should be negligeable in such low HI-column density field as the N1 field ($N_{HI}= 10^{20}$ at/cm^2).

Uncertainties on the magnitudes are smaller than 0.1 mag.

ID	Position (J2000)	P_{ran}	r' Mag.	(U-r')	(U-g')	(g'-r')	K_s mag	Morphology	Comments
FN1-000	160552.8+540651	0.07	17.94			1.46	13.41 ±0.13	Irregular	Merger with tail
FN1-001	160736.6+535730	0.00	16.67	1.28	0.39	0.89	12.29 ±0.09	Sb/c	
FN1-002	161005.8+541029	0.10	17.73	1.6	0.68	0.92	12.63 ±0.11	Sb/c	
FN1-003	161254.1+545526	0.29	17.13	1.94	0.86	1.08	12.57 ±0.09	Sa	Em. with small EW
FN1-004N	161109.5+535808	0.01	16.84	2.36	1.25	1.11	11.80 ±0.07	Sa/b	Star forming knots (Seyfert)
FN1-004S	161107.3+535711	1	19.03	1.59	0.74	0.85		Sa/b	
FN1-005	160443.4+543331	0.84	19.22	2.89	1.41	1.48		2 Spirals	Merger
FN1-006	160433.9+544432	0.32	17.26	1.77	0.83	0.94	12.79 ±0.10	Sa/O	
FN1-007	161331.2+541630	0.03	17.18	1.91	1.02	0.89	12.64 ±0.09	SO/a	
FN1-009	160803.7+545302	0.01	16.58	1.58	0.80	0.78	11.88 ±0.07	Sa	Liner
FN1-011	160808.6+535242	0.13	17.84	1.77	0.81	0.96	13.21 ±0.16	Sa	
FN1-012	161214.0+540833	0.30	17.88	1.38	0.53	0.85	13.69 ±0.18	Sb	
FN1-014	161551.4+541536	0.43	19.20	1.32	0.36	0.96		Sa/b	
FN1-015	160724.7+541212	0.37	18.89	1.53	0.51	1.02	14.8 ±0.1 [a]	Sa/b	
FN1-016	16-738.0+544602	0.97	17.09	1.80	0.82	0.98	13.14 ±0.10	Elliptical	
FN1-018	161407.1+541920	0.87	18.92	2.18	1.08	1.1	13.31 ±0.20	Sa/b	Edge-on (AGN)
FN1-020	160812.3+545526	0.88	19.72	2.62	1.39	1.23		Disk?	Very faint
FN1-021	161308.0+545142	0.33	18.18	1.60	0.65	0.95	13.05 ±0.12	Sa/b	
FN1-021B	161319.2+545137	1	18.04	0.44	-0.10	0.54		E/SO	

[a]: K magnitude from Sajina et al. (2003)

Table 2. Photometric properties and morphology of the FIRBACK fully identified sources.

ID	Position (J2000)	P_{ran}	r' Mag.	(U-r')	(U-g')	(g'-r')	K_s mag	Morphology	Comments
FN1-023	160835.4+535020	0.40	18.94	2.39	1.42	0.97	13.63 ±0.19	Sb/c	
FN1-024	160937.5+541259	0.54	18.36	2.07	0.93	1.14	13.50 ±0.15	SO	
FN1-026	161436.9+541635	0.27	19.10	1.82	0.83	0.99	13.75 ±0.18	Sb	Edge-on, dust lanes
FN1-031	161103.7+544322	0.05	17.69	1.25	0.41	0.84	13.36 ±0.16	SO/a ?	
FN1-033	161301.1+541003	0.22	18.47	1.40	0.41	0.99		E/SO	
FN1-035	161527.7+543414	0.39	18.11	1.65	0.67	0.98	13.38 ±0.14	Sa/b	
FN1-038	160747.3+534836	0.53	19.26	1.85	0.93	0.92		Sa/b	
FN1-039	160848.9+545151	0.59	20.18	1.90	0.71	1.19	15.8 ±0.2 [a]	?	Possibly disk
FN1-041	160814.2+542836	0.30	18.89	1.19	0.36	0.83	14.7 ±0.1 [a]	SO/a	
FN1-043	160553.4+542226	0.59	18.69	1.74	0.54	1.20	14.00 ±0.20	Sb	Barred

[a]: K magnitude from Sajina et al. (2003)

Table 2. Photometric properties (continued)...

ID	Position (J2000)	P_{ran}	r' Mag.	(U-r')	(U-g')	(g'-r')	K_s mag	Morphology	Comments
FN1-008	160858.0+541818	0.21	18.83	2.25	0.81	1.44	13.76 ±0.19	SO/a ?	Two objects
FN1-010	160930.9+535148	0.26	17.18	2.09	1.04	1.05	12.63 ±0.14	Sa/b	Prob. AGN
FN1-019	161235.4+541545	0.01	17.05	1.95	0.94	1.01	12.31 ±0.10	Sa/b	AGN
FN1-025N	160833.4+545510	0.88	18.81	1.43	0.64	0.79		Sb/c	
FN1-025S	160834.5+545421	??						Sb	
FN1-028A	160738.2+534250	0.47	17.59	2.04	1.05	0.99	13.68 ±0.18	SO/a	
FN1-028B	160746.5+534151	1	16.70	2.20	1.18	1.02	12.26 ±0.09	SO	
FN1-029	161117.5+541629	0.26	18.67	2.38	1.38	1.00	13.65 ±0.18	Sa/b	
FN1-032	161239.9+543657	0.07	17.38	2.33	1.34	0.99	12.07 ±0.08	Sa	AGN
FN1-034	160724.2+544330	0.33	18.66	0.90	0.19	0.71	19.3 ±0.8[a]	Sa/b	
FN1-037	161509.0+541837	0.36	19.42	1.88	1.14	0.74		Sb	Id. uncertain
FN1-042	161038.9+543628	0.08	17.24	2.24	1.35	0.89	12.06 ±0.08	Sb	AGN
FN1-044	160931.6+541827	0.74	18.69	1.99	0.93	1.06		Sa	
FN1-045	160853.9+544735	0.6	19.24	2.12	1.11	1.01	14.33 ±0.17	Sa	
FN1-046A	161249.4+540837	1	18.01	2.44	1.22	1.22	13.19 ±0.13	SO/a	Radio source
FN1-046B	161251.1+540801	0.52	17.38	1.61	0.68	0.93	13.09 ±0.12	SBa	Barred
FN1-048	161059.1+542305	0.64	19.37	1.31	0.44	0.87	19.3 ±0.9[a]	Sb/c	
FN1-049	160801.6+543642	0.54	18.89	1.54	0.69	0.85	13.73 ±0.20		
FN1-057	160801.6+543643	0.78	18.43	1.64	0.69	0.95	13.26 ±0.15	SBa	Barred
FN1-101	160946.1+542127	0.11	18.44	0.85	0.25	0.60		Sa/b	

[a]: K magnitude from Sajina et al. (2003)

Table 3. Photometry of additional sources (sources with multiple or uncomplete identification); or sources outside the 4σ sample (FN1-057 and FN1-101).

4.2. Spectral properties

Most of the spectra display both emission and absorption lines. All lines have been used to derive the redshift (except those with low signal to noise ratio, indicated in brackets in Table 4 and in Table 5). The 1σ dispersion is given after the velocity: it should be smaller when more lines are available, but the well-known systematic difference between emission and absorption lines also affects the result when both types of lines are used. To this internal error has to be added the (random) external error, which should not exceed one pixel after correction for flexures, so that the final uncertainty in velocities should be better than 150 km/s in most cases. The distribution in redshifts of the identified sample is plotted in Fig. 1. Out of the 28 identified sources, about 80% have redshifts lower than 0.25. This is in very good agreement with the redshift distribution of the Lagache et al. (2003) model of evolution of IR galaxies.

The emission line spectra are very similar to those of standard IRAS galaxies (e.g. Veilleux et al. (1995)), with the classical Balmer lines (Hα and Hβ essentially) and forbidden lines of oxygen (3727, 5007, and, less often, 6300$\AA$), nitrogen (6548, 6584$\AA$) and sulfur (6717, 6731$\AA$). In most cases, the excitation is low, and the continuum appears to be reddened, so that quite often the Hβ or [OIII] lines are not even detected. Because the Hβ emission, when detected, is often superposed on the corresponding absorption line, the correct estimate of its intensity would require a proper determination (and subtraction) of the underlying absorption, which was not possible with the available spectra. We therefore have not derived an estimate of the reddening from the Hα/Hβ ratio. This prevents us from deriving reliable Hα luminosities, even for those objects observed under photometric conditions: the values given are therefore only lower limits.

We have not detected any broad Balmer line objects in this sample. The signal to noise ratio in the continuum would however not have allowed us to detect faint, broad wings in most cases (for Seyfert types 1.9 or 1.8). But we have a number of objects where the [NII]/Hα ratio is larger than the usual HII region limit of 0.5 (e.g. Veilleux (2002)), and so is the [SII]/Hα ratio, therefore indicating the presence of an AGN. Only exceptionnaly are other line ratios available, preventing a more precise classification of this AGN: the usual weakness or absence of the [OIII] line however points preferentially towards a Liner rather than a Seyfert 2 galaxy. We have 8 such cases out of a total of 50 objects, that is 15 %, a proportion similar to the one found in other IR-selected samples of low redshift objects with moderate IR luminosities.

ID	Coordinates[a]	cz (km/s)	r' Mag.	Emission (Hα flux)	Absorption	Comments
FN1-000	160552.8+540651	43604 ±87	17.94	Hα(183), [NII], [OI], [NI], Hβ, [OIII] Hγ, [OII], (HeI)	Na	Liner
FN1-001	160736.6+535730	9004 ±51	16.67	Hα(93.9), [NII], [SII], Hβ, [OIII], HeI	Na, Ca, (Mg), G, H, K	
FN1-002	161005.8+541029	19119 ±119	17.73	Hα(13.9), [NII], [SII], Hβ, [OII]	Na, Mg, G, H, K	
FN1-003	161254.1+545526	19481 ±90	17.13	Hα(21.7), [NII], [SII], [OII]	Na, (Hβ, Mg, Ca, G, H, K	Em. with small EW
FN1-004N	161109.5+535807	19107 ±85	16.84	Hα(9.5), [NII], [SII], [OIII]	Na, Hβ, G, Mg	Sey2, [NII]/Hα>1
FN1-004S	161107.2+535711	19208 ±88	19.03	Hα(5.5), [NII], [SII]	Na, G	
FN1-005	160443.4+543331	86400[b]	19.22			
FN1-006	160433.9+544432	22608 ±63	17.26	Hα(34.8), [NII],[SII], [OII]	Na, Mg, Hβ, G, Hδ, H, K	
FN1-007	161331.2+541630	18435 ±41	17.18	Hα(59.1), [NII], [SII], [OII]	Na, G, H, K	
FN1-009	160803.7+545302	15746 ±865	16.58	Hα(17.0), [NII], [SII] [OI], [OIII], [OII]	Na, Mg, Ca, Hβ, Hγ, Hδ G, H, K	Liner
FN1-011	160808.6+535242	19326 ±48	17.84	Hα, [NII], [SII], Hβ	Na, Ca, G, (H), K	
FN1-012	161214.0+540833	19770 ±73	17.88	Hα(32.7), [NII], [SII], Hβ, [OIII], [OII]	Na, Mg, Ca, H, K	
FN1-014	161551.4+541536	21000[c]	19.20			
FN1-015	160724.7+541212	70183 ±58	18.89	Hβ, [OII]	H, K	
FN1-016	16-738.0+544602	27579 ±85	17.09	Hα(48.7), [NII], [OII]	Na, (Hβ), (Hγ), Hδ, H, K	
FN1-018	161407.1+541920	25598 ±105	18.92	Hα(7.9), [NII], ([OIII]), [OII]	Na	AGN ([NII]/Hα~1)
FN1-020	160812.3+545526	3000[c]	19.72			
FN1-021	161308.0+545142	19349 ±39	18.18	Hα, [NII], [SII], Hβ	Na	
FN1-021B	161319.2+545138	19983 ±18	18.04	Hα, [NII], [SII], Hβ, [OIII]		

[a]: J2000 position of the optical counterpart. [b]: Chapman (private communication) [c]: photometric estimate (Babbedge et al. (2004))

Table 4. Spectroscopic properties of the FIRBACK identified sources. The quoted uncertainty for velocities is the internal (1σ) error, and does not include the systematics (see text). Emission (or absorption) lines shown in brackets are seen, but not used for the velocity determination because of poor S/N. The Hα flux is given, when conditions were photometric, in units of $10^{-16} ergs/cm^2/s$.

ID	Coordinates	cz (km/s)	r' Mag.	Emission (Hα flux)	Absorption	Comments
FN1-023	160835.4+535020	18780 ±26	18.94	Hα(26.1), [NII], [SII], Hβ, [OIII], [OII]	Na	
FN1-024	160937.5+541259	25770 ±60	18.36	Hα(20.1), [NII], Hβ, ([OII])	Na	
FN1-026	161436.9+541635	23835 ±12	19.10	Hα(4.8), [NII]	(Na), (Hδ)	
FN1-031	161103.7+544322	18840 ±78	17.69	Hα(8.6), [NII], [SII], ([OIII])	H, K	
FN1-033	161301.1+541003	41173 ±65	18.47	Hβ, [OII]	Na, G, H, (K)	
FN1-035	161527.7+543414	40820 ±72	18.11	Hα, [NII], [OII]	Na, Mg, (Hβ), H, K	
FN1-038	160747.3+534836	32192 ±20	19.26	Hα(10.1), [NII], [SII], Hβ, [OIII]		
FN1-039	160848.9+545151	80700[b]	20.18			
FN1-041	160814.2+542836	35995 ±46	18.89	Hα, [NII], Hβ, [OIII], Hγ		
FN1-043	160553.4+542226	63470 ±53	18.69	Hα, [NII], [OII]	(G), H	

Table 4. Spectroscopic properties (continued)... For source FN1-040, $z = 0.45$, see Chapman et al. (2002)

ID	Coordinates [a]	cz (km/s)	r' Mag.	Emission (Hα flux)	Absorption	Comments
FN1-008	160858.0+541818	78770 ±83	18.83	[OII]	Na, H, K	
FN1-010	160930.9+535148	18741 ±75	17.18	Hα(9.1), [NII], [SII]	Na, Mg, Ca, Hβ, G, H, K	Prob. AGN (NII/Hα>0.5)
FN1-019	161235.4+541545	18663 ±104	17.05	Hα(5.0), [NII], [OI], [OII]	Na, (Mg), Hβ, G, H, K	AGN ([NII]/Hα>0.5
FN1-025N	160833.4+545510	15761 ±37	18.81	Hα, [NII], [SII], Hβ, [OIII], ([OII])	H	
FN1-025S	160834.5+545421	32671 ±54		Hα, [NII], Hβ, [OIII], [OII]		
FN1-028A	160738.2+534250	19816 ±154	17.59		Na, Mg, Hβ, G, H, K	
FN1-028B	160746.5+534151	19921 ±142	16.70	[OII]	Mg, Ca, (Hβ), G, H, K	
FN1-029	161117.5+541629	43100 ±70	18.67	Hα(11.4), [NII]	Na, H, K	
FN1-032	161239.9+543657	18674 ±63	17.38	Hα(8.9), [NII], [SII]	Na, Mg, (Hβ), G, H, K	AGN (NII/Hα~1)
FN1-034	(160724.2+544330)	27916 ±54	18.66	Hα(27.5), [NII], Hβ, [OIII], [OII]	K	
FN1-037	161509.0+541837	85500 ±300	19.42	[OII], [NeV]	Mg, K	Id. uncertain
FN1-042	161038.9+543628	18844 ±68	17.24	[NII]	Na, Hβ, Ca, Mg, H, K	AGN ([NII]/Hα>1)
FN1-044	160931.6+541827	24682 ±41	18.69	Hα(7.2), [NII], [SII], [OIII]	(Na), (G)	
FN1-045	160853.9+544735	276400?	19.24	[OII] ??		Single emission line
FN1-046A	161249.4+540837	45290 ±16	18.01	([NeV]??)	Na, Mg, (G), H, K	Radio source
FN1-046B	161251.1+540801	18566 ±73	17.38	Hα, [NII], [SII], [OII]	Na, G, H, K	
FN1-048	161059.1+542305	18688 ±22	19.37	Hα(10.8), [NII], [SII], Hβ, [OIII]	Na	
FN1-049	160801.6+543642	38381 ±107	18.89	Hβ, [OIII], [OII]	(H), K	
FN1-057	160801.6+543643	27546 ±35	18.43	Hα(15.1), [NII]	Na, Ca, (G), H, K	
FN1-101	160946.1+542127	19690 ±58	18.44	Hα, [NII], [SII], Hβ, [OIII], [OII]	H, K	

[a]: J2000 position of the optical counterpart.

Table 5. Spectroscopy of additional sources (sources with multiple or uncomplete identification; or sources outside the 4σ sample (FN1-057 and FN1-101)). The Hα flux is given in parenthesis, in units of $10^{-16} ergs/cm^2/s$, when conditions were photometric.

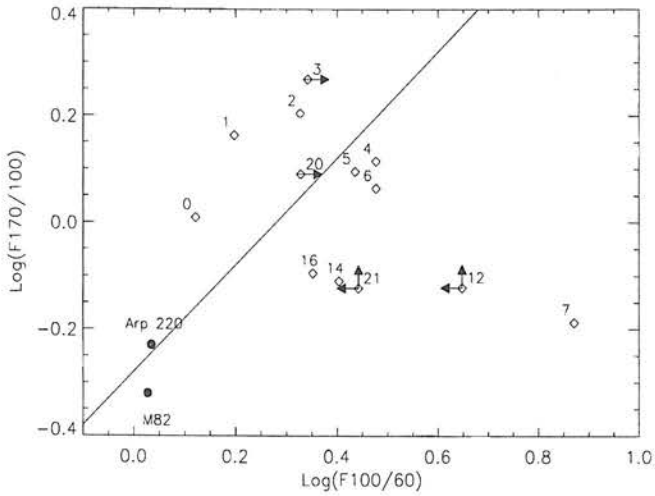

Fig. 3. 60, 100, 170 color ratios. The continuous line is the ISS fit, see text.

5. Mid and far-IR properties

5.1. For sources with IRAS detections: 60, 100 and 170 colors

The mean infrared (60, 100 and 170 μm) colors have been calculated for our identified galaxies. The two-color diagram $\log(F_{170\mu m}/F_{100\mu m})$ versus $\log(F_{100\mu m}/F_{60\mu m})$ (Fig. 3) includes all galaxies with measurements or upper limits in the 60 and 100 μm IRAS bands . The continous line has a slope of ≈ 1 and corresponds to the least square bisector regression for the ISOPHOT 170 μm serendipity survey (ISS, Stickel et al. (2000)). Most of the galaxies have a behavior that is consistent with those from the ISS, apart FN1-007 (in the lower right corner of the figure; this object is peculiar, see the discussion of SEDs later on). Their colors are well within the sequence of the normal galaxy sample of Dale et al. (2001), and correspond more to the quiescent end of their classification (i.e. the coldest galaxies). Note that for Arp220 and M82, $\log(F_{100\mu m}/F_{60\mu m}) \approx 0.03$, and $\log(F_{170\mu m}/F_{100\mu m}) \approx 0.25$, ratios that are more typical of active galaxies.

5.2. The 15/170 μm color

Twenty four galaxies from our identified sample have both $F_{170\mu m}$ and $F_{15\mu m}$ measurements. We add to those sources $FN1-013$ and $FN1-053$ which do not yet have a redshift measurement but have a clearly identified 15 μm counterpart. From their $F_{170\mu m}/F_{15\mu m}$ ratio, the contribution of these 26 galaxies to the CIB can be estimated. The observed color of the CIB can be bracketted from the work of Elbaz et al. (2002) and Renault et al. (2001) : $30 < CIB_{170\mu m/15\mu m} < 75$. On the other hand, the model of Lagache et al. (2004), which reproduces (i) the number counts at 15, 24, 60, 70, 90, 160, 170 and 850 μm, (ii) the known redshift distributions (mainly at 15, 170

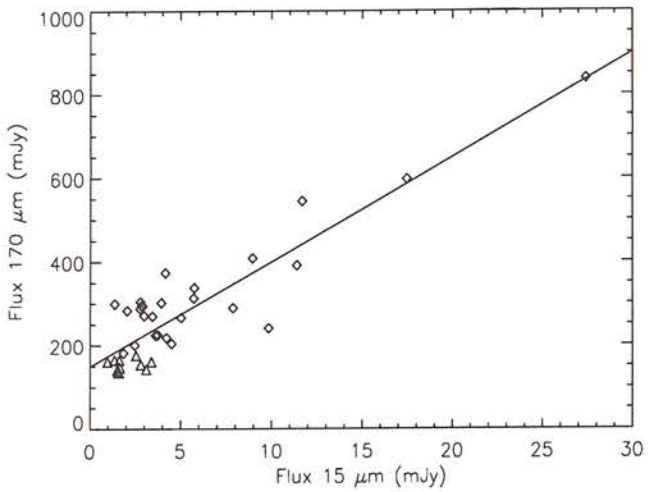

Fig. 4. 170/15 correlation. The diamonds are for $S_{170} > 4\sigma$ and the triangle for $3\sigma < S_{170} < 4\sigma$. The continuous line is the best fit for all sources. It has a slope of 170/15=25.

and 850 μm), (iii) the local luminosity functions at 60 and 850 μm and (iv) the CIB (from 100 to 1000 μm) and its fluctuations (at 60, 100 and 170 μm), gives a ratio $CIB_{170\mu m/15\mu m} = 59.4$. In view of the excellent agreement of the Lagache et al. (2004) model with the observations from the mid-IR to the mm range, and the fact that the CIB color from the model lies well within the range given by the observations, we take this value from the model (60) as the best estimate of the 170/15 color of the CIB.

The 170/15 relation for the observed FIRBACK galaxies is shown in Fig 4. For sources with $S_{170} > 4\sigma$, the slope of the correlation is equal to 22.3. If we add to this sample the sources with $3\sigma < S_{170} < 4\sigma$, which are not identified optically, but have clear 15 μm counterparts, the slope increases from 22.3 to 25, but is still more than a factor of 2 below the mean color of the CIB as discussed above. These sources are thus clearly not representative of the bulk of the sources contributing to the 170 μm CIB. This is not surprising, as (1) it was shown by Dole et al. 2001 that the bright FIRBACK sources (down to the 3σ limit of the survey) represent in practice less than 5% of the CIB at 170 μm and (2) the observed CIB 170 μm/15 μm ratio is about three times higher than that of the galaxies studied here, and is better matched with higher z sources due to the K-correction (around z $\sim$ 0.8, as observed by Elbaz et al. (2002)).

5.3. Spectral Energy Distributions

5.3.1. Comparison with template spectra

To derive an IR luminosity when only a few sampling points are available, the most reliable technique is to identify a template spectrum whose Spectral Energy Distribution

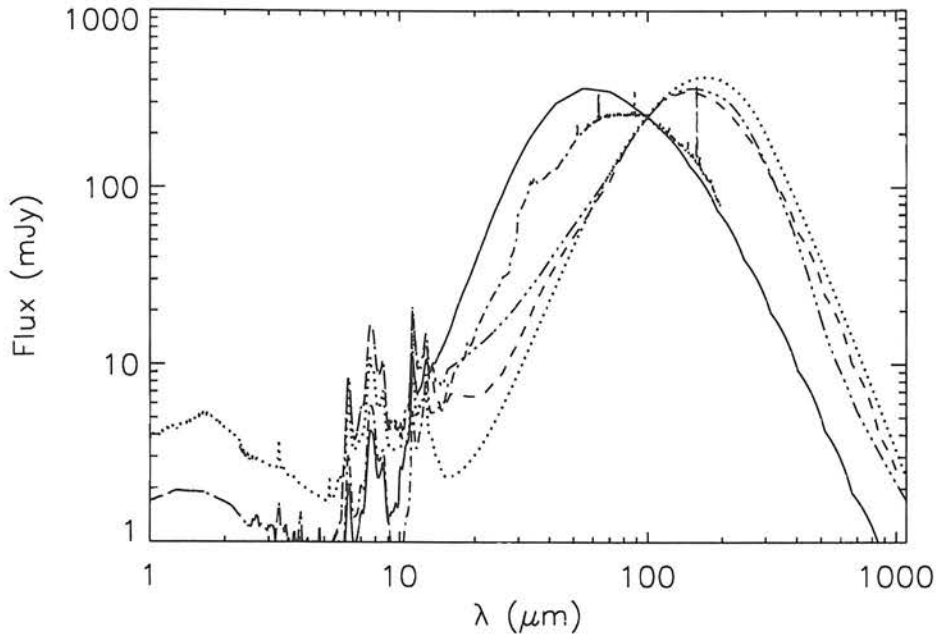

Fig. 5. Comparison of template SEDs from Dale et al. (2001) (α=2.6, quiescent case, dashed line; and α=1, active case, continuous line) and Lagache et al. (2003) ("cold galaxy", triple-dot-long-dashed line) with observed SED's. The dotted line is the observed SED of M101, the dashed-dotted line the SED of M82 (curves normalised at 100 μm).

(SED) best matches the observations (although the solution might not be unique). There are plenty of SEDs that could be tried: the ones we used are shown in Fig. 5 and are discussed below. Our first tests have shown (see also Patris et al. (2003)) that typical starbursts SEDs like (M82) did not match the data, and that colder components were required. We thus used template spectra from Lagache et al. (2003), most notably their "normal-galaxy" spectrum, that is derived from observations of a sample of galaxies with measurements from 15 to 850 μm (it was built essentially from cold galaxies, including those from the present sample, and will therefore be called "the cold template" in what follows, to avoid confusion). This"cold-galaxy" template has been fitted to each of our objects, and is shown as a solid line superposed on the observed points in Fig. 6 and Fig. 7. The SED of the typical starburst galaxy M82 is also shown as a dashed line. The important point to note is that the M82 template, although fitting quite well the 170/15 color, fails to reproduce the spectra at 60 and 100 μm. This template, although widely used, is not appropriate for the FIRBACK sources, which are colder galaxies.

For a quantitative appraisal of the contribution of various components, we use the models of Dale et al. (2001, 2002): they have modeled the infrared spectral energy distribution of normal star-forming galaxies, as a function of a parameter α which quantifies the relative contribution of "active" and quiescent regions from galaxy to

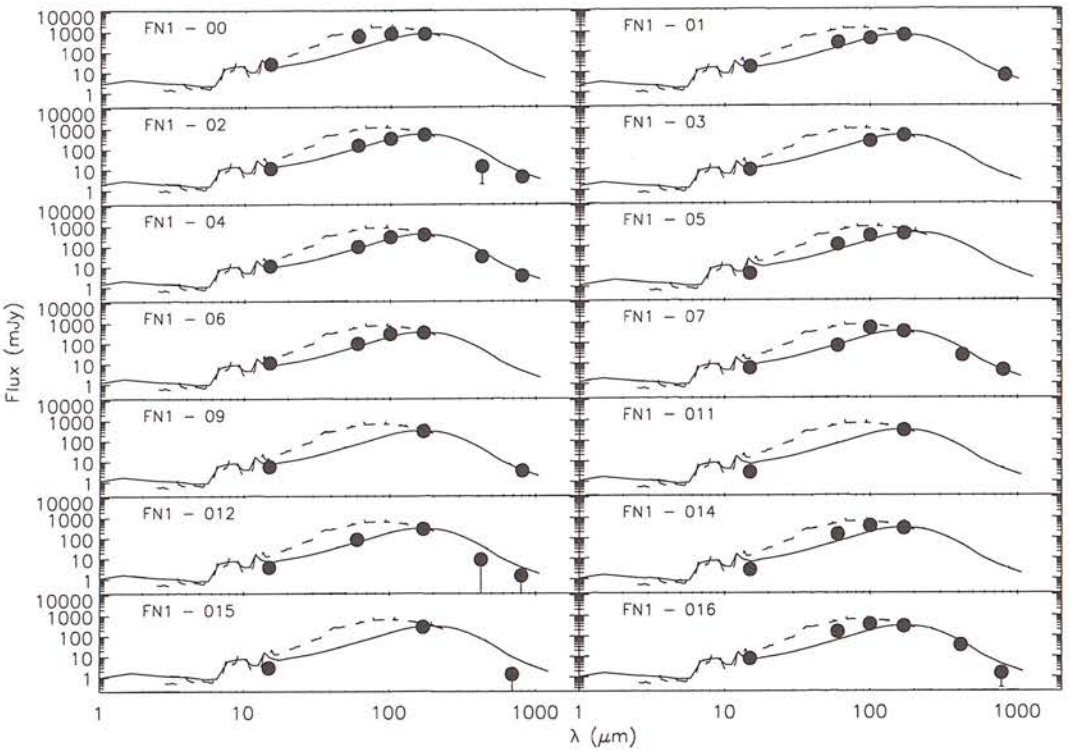

Fig. 6. The results from the fit of the source SEDs with a "cold galaxy" spectrum from Lagache et al. (2003). The continuous line is the "cold galaxy" spectrum template, the dashed line the spectrum of M82.

galaxy, $\alpha \approx 2.6$ being at the quiescent end and $\alpha \approx 1$ at the active end. Nine of our identified galaxies have both $F_{60\mu m}$ and $F_{100\mu m}$ measurements available and all but $FN1-007$ have a $(F_{60\mu m}/F_{100\mu m})$ ratio that falls in the range of colors modeled by Dale et al. (2001, 2002). From the observed $(F_{60\mu m}/F_{100\mu m})$ ratio, we derive α, and the corresponding best fit template spectrum. The best fit template spectra from Dale et al. (2002) are shown in Fig. 8. The derived α (Table 6) corresponds always to models where the quiescent component dominates ($\alpha > 1.6$): this clearly shows that our sample is composed of preferentially cold galaxies with only moderate star formation, rather than more active objects where the $F_{60\mu m}/F_{100\mu m}$ is usually greater than 1.

The chi squared of the two fits (the Dale et al. set or the Lagache et al. "cold galaxy" spectrum; the M82 template is clearly not adequate, as can be seen from the figures) have been compared. The Dale et al. template spectrum is a better fit for only FN-000 and FN-016, while it is comparable to the "cold galaxy" spectrum fit for the other galaxies. The FN-007 object remains an exception, but its optical spectrum does not give any indication of a peculiar nature. Its SED is however peculiar, with a very high $F_{100\mu m}/F_{60\mu m}$ and a comparatively low $F_{170\mu m}/F_{100\mu m}$ ratio, difficult to explain with

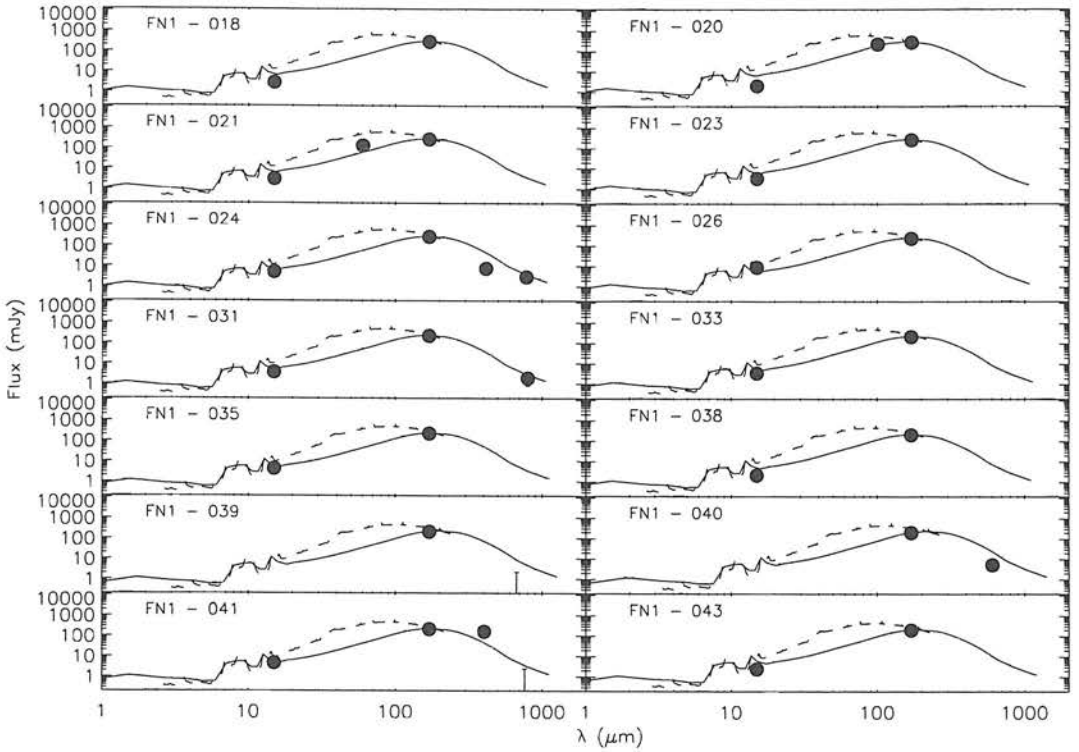

Fig. 7. Same as in Fig 6 for the remaining identified sources.

standard components. On average, the FIRBACK sources are clearly cold galaxies.

5.3.2. IR luminosities

Total luminosities have then been calculated in several ways. First by integrating the best fit spectra between 1 and 1000 μm. This was done for all sources with the "cold" spectrum, even for those where only one measured point was available, namely the 170 μm flux. The derived values are shown in Table 6. For those objects where a "Dale" spectrum could be determined, the luminosity is also derived (Table 6). Note that this second method was not applied to FN1-007, although it has measured 60 μm and 100 μm fluxes, because its color lies outside the range of validity of Dale's models. Finally, we used also the formula proposed by Stickel et al. (2000) to calculate luminosities when fluxes at 60, 100 and 170 μm were available. Following their formula :

$$F_{40-220} = 1.34 \times 10^{-14} \times \quad (2.58 F_{60\mu m} + 1.00 F_{100\mu m} + 0.63 F_{170\mu m}) \quad \mathrm{Wm}^{-2} \qquad (1)$$

$$F_{1-1000} = 1.35 F_{40-220} \quad \mathrm{Wm}^{-2} \qquad (2)$$

$$L_{1-1000} = 4\pi D^2 F_{1-1000} \qquad (3)$$

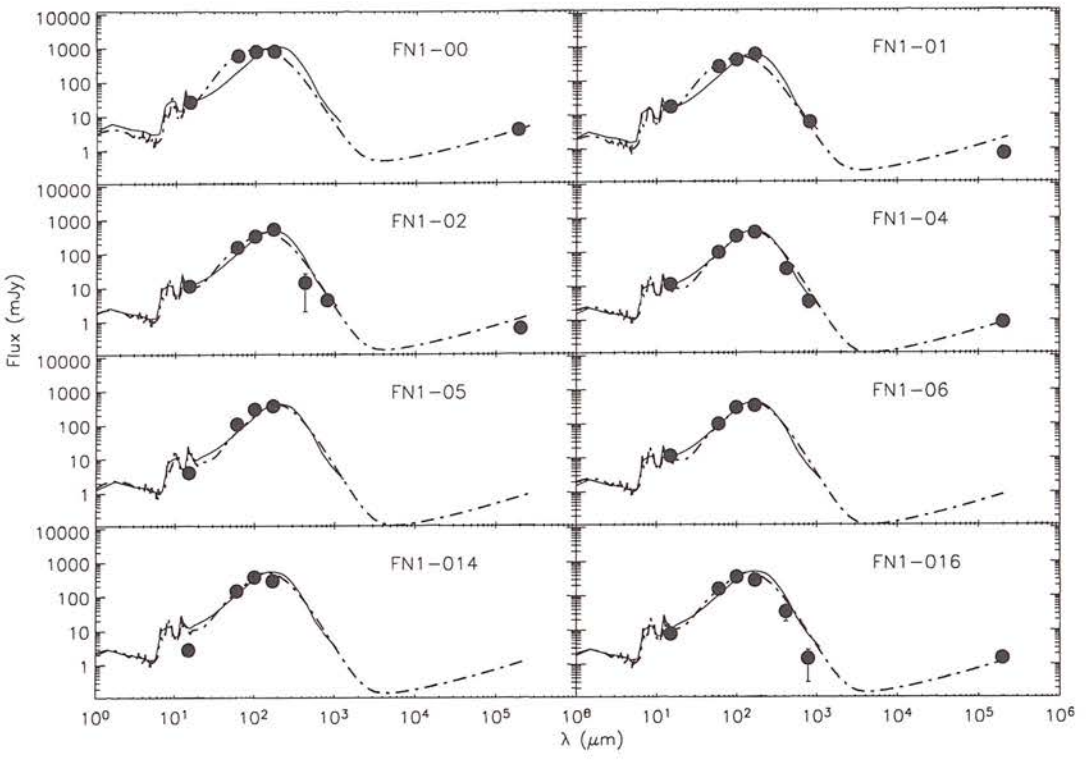

Fig. 8. SEDs of those identified sources that have both measured 60 μm and 100 μm fluxes are fitted with a Dale model. The dashed-dotted line is the Dale model fit, the thin continuous one the "cold galaxy" spectrum template (Lagache et al. (2003)). Both the model and the template have been normalized at 170 μm.

where D is the luminosity distance. The luminosities derived with this formula are also shown in Table 6.

For sources for which 60 and 100 μm fluxes are available, the estimated luminosities can be compared. The luminosities estimated from Dale's best fit model are generally higher than the luminosities derived by the two other methods: this can be attributed to the fact that Dale's models better take into account the contribution from the cold emission at very long wavelengths. But the differences are always smaller than a factor of two, and generally lower than 30 %. In the subsequent analysis, we shall use (unless otherwise stated) the luminosity calculated from the Lagache model, as it is available for all objects.

6. The radio-FIR relation

The far-IR (FIR)-radio relation was introduced by Helou et al. (1985) from the study of normal galaxies with IRAS, and was shown to be quite general for various samples of galaxies (see Condon (1992) for a review). The tightness of the relation, and its small

ID	z	Lagache Lum. ($10^{10} L_\odot$)	Stickel L. ($10^{10} L_\odot$)	Dale L. (and α) ($10^{10} L_\odot$)	Hα L. ($10^{6} L_\odot$)	K Lum. ($10^{9} L_\odot$)	SFR(IR) $M_\odot/yr$	SFR(Hα) $10^{-2} M_\odot/yr$	q
0	0.145	52.3	73.0	95.3 (1.69)	222	5.87	68	>675	2.30
1	0.030	1.5	1.3	1.8 (1.83)	4.9	0.70	1.9	>15	2.69
2	0.063	6.4	4.6	6.1 (2.07)	3.2	2.32	8.3	>9.7	2.60
3	0.063	4.9			5.2	2.54	6.4	>16	
4	0.063	4.6	3.4	5.2 (2.52)	2.2	4.97	6.0	>6.7	2.32
5	0.288	100.8	70.5	113.6 (2.35)			131		
6	0.073	5.6	4.4	7.2 (2.52)	11.3	2.80	7.3	>34	
7	0.063	3.6	3.8		12.8	2.14	4.7	>39	2.35
9	0.053	2.4			2.8	3.14	3.1	>8.5	
11	0.063	3.6			3.0	1.39	4.7	>9.1	
12	0.063	3.8			8.2	0.93	4.9	>25	2.66
14	0.07	4.1	4.8	7.9 (2.25)			5.3		
15	0.23	48.6				4.23	63		
16	0.090	6.8	8.5	13.3 (2.13)	23.6	3.02	8.8	>72	2.22
18	0.067	6.0			3.3	2.22	7.8	>10	
20	0.01	0.1					0.13		
21	0.064	3.2			3.6	1.61	4.1	>11	
23	0.063	3.1			5.9	0.89	4.0	>18	
24	0.086	5.7			8.5	1.89	7.4	>26	
26	0.077	4.3			1.7	1.28	5.6	>5.2	
31	0.063	2.6			1.9	1.15	3.4	>5.8	
33	0.133	12.4					16.1		
35	0.137	11.9			32.5	5.30	15.5	>99	
38	0.107	6.9			6.7		9.0	>20	
39	0.269	47.6				2.23	62		
40[a]	0.45	122.9				0.17	160	2000	
41	0.12	8.6				1.22	11.1		
43	0.21	27.5			21.1	7.23	36	>64	
49	0.127	9.8				3.39	12.7		

[a]: Data from Chapman et al. 2002, main component, reddening corrected Hβ luminosity.

Table 6. Source luminosities calculated from the integrated "normal-galaxy" spectrum (Lagache Lum.), Stickel's formula (Stickel Lum.) and the best fit Dale spectrum (Dale Lum.) with its α value. The Hα luminosities, uncorrected for extinction (hence the ">" sign in the corresponding SFR), are also given. The K luminosity is directly derived from the K magnitude. The "q" parameter is the logarithmic ratio of IR to radio flux, following Helou et al. (1985).

dispersion, is best represented by the "q" parameter introduced by Helou et al. (1985), which is the ratio of FIR (as measured by IRAS) to the radio continuum (1.4 GHz) flux densities:

$$q = log[(FIR/3.75 \times 10^{12} \quad \mathrm{Wm^{-2}})/(\mathrm{S}_\nu/\mathrm{Wm^{-2}Hz^{-1}})] \tag{4}$$

where FIR is the 40-120 μm flux density derived from IRAS measurements via:

$$FIR/Wm^{-2} = 1.26 \times (2.58 S_{60\mu m} + S_{100\mu m}) \tag{5}$$

The average value for q was found to be around 2.3 with a scatter lower than 0.2 (Helou et al. (1985)).

We can calculate this parameter for those FIRBACK objects where IRAS data are available, and found similar values (Table 6): the objects lie slightly above the mean relation, with an average of 2.45 for 8 objects. To increase the number of objects, one could calculate backwards a FIR from the integrated IR luminosity (for instance from the "Lagache" luminosity which is available for all of them, see previous section): this would then represent a FIR_{1-1000} instead of FIR_{40-120} and, as expected, the q ratio is increased accordingly. But only few objects are added this way (due to the limited radio detections), so that no advantage is obtained in practice from this different approach.

The main result is that no object has a q value significantly lower than the average 2.3 found in normal galaxies: there is therefore no evidence for a significant contribution to the total energy balance from a radio-loud AGN in any of those objects. For those few objects where the presence of an AGN is indicated from the spectroscopic data, and where radio data are also available, the q ratio is not higher, suggesting that the contribution of the AGN to the FIR flux is also negligible. However, we cannot exclude a contribution from a radio-quite AGN.

7. Star formation rates

Several indicators can be used to estimate the Star Formation Rate (SFR) of galaxies, but it is widely accepted that the far-IR luminosity gives the best estimate, because it measures the bolometric luminosity of the object: this is correct as long as one can be sure that a hidden AGN is not significantly contributing to the total energy output. We have used the IR luminosity computed from the fit of the "Lagache" template (which is available for all the identified objects) and derived the SFR with the calibration of Devriendt et al. (1999):

$$\mathrm{SFR\,(M_\odot per\ year)} = \frac{\mathrm{L_{IR}\,(10^9 L_\odot)}}{7.7}$$

The result is given in Table 6. Most of the objects have star formation rates of only a few $M_\odot$/year and are therefore moderate starbursters only. Five objects are

however standing out: $FN1-000$, $FN1-005$, $FN1-015$, $FN1-039$, and $FN1-040$. $FN1-005$ and $FN1-040$ are the most distant objects in this limited sample of fully identified sources, and are clearly ULIRG's detected at larger distance than the average, bright, FIRBACK sources (for details on $FN1-040$, see Chapman et al. 2002). The same is true for $FN1-015$ although it has a three times smaller IR luminosity. For $FN1-039$ we have no detailed information. More interesting is the case of $FN1-000$, which is closer, and at similar distances than, for instance, $FN1-035$ or $FN1-033$, but has a five times greater luminosity. We have no evidence that an AGN is located inside this galaxy neither from spectroscopy nor from the q parameter, so we consider it as a vigorous, local starburst galaxy.

We can, in principle, also use the Hα luminosity to derive the SFR, following the recipe of, e.g., Kennicutt (1992):

$$\mathrm{SFR\ (M_\odot per\ year)} = \frac{\mathrm{L_{H\alpha}\ (10^{41} erg/s)}}{1.26}$$

The result is also given in Table 6. For the 17 objects where we have both IR and Hα estimates, the latter always gives a much lower SFR, on average a factor of 22 lower than the former one (note that the 4 high-luminosity objects are not included in this sample). This is clearly due to the fact that our Hα fluxes are not corrected for internal extinction. We have no reddening estimate for each individual galaxy. In a few cases only, like $FN1-000$, an extinction can be estimated from the Hα over Hβ ratio (where we find $A_v = 2$, with the necessary caveat due to the unresolved underlying absorption lines). We could apply to the whole sample an average extinction correction of $A_v \sim 1$, as was done in other cases (e.g. Pettini et al. (2001) where a low extinction was estimated from the comparison between UV and visible properties or Lilly et al. (2003) where this value was derived from a local reference sample). It is however very unlikely that such a low extinction would be appropriate for our sample which is mainly emitting in the far IR. Instead, we can use a value obtained from a very similar sample where more data are available, namely the FIRBACK southern field (Patris et al. (2003)): the average extinction found there is $A_v = 3$. This corresponds to an upwards correction of a factor of 10 to the SFR derived from the Hα luminosities. While such a correction is only indicative, because the dispersion of extinction values from object to object may be very high, the similarity of these two samples ensures that the selection effects are minimized: this value is certainly more appropriate than the standard $A_v = 1$ value applied in other samples. After such correction, it appears that the SFR derived from Hα becomes closer to the one derived from the IR, although still systematically lower. The remaining difference is due in part to some slit losses and, more probably, to a large optical thickness in Hα thatprevents an accurate extinction correction. This problem is

similar to the one encountered in other IR selected samples (like in Patris et al. (2003)) but without additionnal data, the comparison between the various SFR estimators cannot be pursued further.

The radio-FIR relation, discussed in the previous section, shows however that the "q" parameter for these galaxies (at least the 8 for which it could be determined) is not significantly different from the standard value found by Condon & Broderick (1991) in the Bright Galaxy Sample of Soifer et al. (1989): 2.4 for our objects instead of 2.3 in the BGS. Our galaxies, with the exception of the 5 cases mentioned above, are thus effectively moderate starbursters only.

8. Discussion and conclusions

Out of the 56 FIRBACK N1 sources with 170 μm fluxes above the 4σ limit of 180 mJy, 28 have been firmly identified and 17 others have at least one of the contributors identified. Only 11 sources remain unclear, essentially due to the lack of additional data (radio or near-IR) with better positional accuracy. Most of the identified sources are quiescent star-forming galaxies which exhibit a colder spectrum than a standard IRAS starburst galaxy. There is no overlap with the catalogue of galaxies resulting from the ISOPHOT 170 μm Serendipity Survey (Stickel et al. 2004), although in view of their detection limit of $\sim$ $0.5Jy$ one would have expected that the first 3 objects from the FIRBACK-N1 catalogue could have been found there. We can nevertheless compare the properties of our objects that have also detected IRAS fluxes with those from the Serendipity Survey discussed by Stickel et al. (2000). Both samples have similar far-IR colours, with an $F_{100\mu m}/F_{60\mu m}$ ratio greater than $\sim$ 1.5 and a $F_{170\mu m}/F_{100\mu m}$ greater than 0.7, which we qualify as "cold" galaxies, dominated by a cold dust component with temperatures roughly between 20 and 40K. This is clearly different from the "warm" LIRGs and ULIRGs, whose $F_{100\mu m}/F_{60\mu m}$ ratio is generally close to or smaller than 1, and which are strong starbursters. The bright FIRBACK galaxies seem therefore to represent a fainter version of the "cold" galaxies detected in the Serendipity Survey.

In terms of optical spectral properties, they resemble the faint IRAS galaxies selected by Bertin et al. (1997) and described in Patris et al. (2003): emission lines, moderate to strong reddening and moderate SFR. There is unfortunately no IR colour information available for the Bertin et al. (1997) sample, for comparison, as those galaxies represent the faintest objects detected in the IRAS survey, in the most sensitive channel ($F_{60\mu m}$ between 150 and 250 mJy), and are thus not detected in the other channels. Its seems however clear, from the three different samples just discussed, that a large population of "cold" galaxies exists at least in the local universe (with a few tens of objects per square degree), with moderate star formation, whose contribution to the global star formation rate is probably significant, although it has been rather neglected until now,

in the "rush" to find always more extreme starburst objects.

The bright FIRBACK galaxies studied here are mostly nearby, Luminous IR Galaxies (LIRGs) with moderate star-formation rates (about 5 to 10 $M_\odot$/yr) and are not particularly associated with detectable merging or interacting systems (only three mergers are clearly detected among the 28 fully identified objects). One could question whether they should be called "starburts galaxies" or whether they do not simply represent disk galaxies with more quiescent star formation (or disks ionised by diffuse radiation). Recent Spitzer observations of classical spiral galaxies like the Scd spiral NGC 300 (Helou et al. 2004) or or the Sc M33 (Hinz et al. 2004) show that their integrated SFR is low, of the order of 0.1-0.2$M_\odot$/yr, and that, although their 160μm emission might be more diffuse, the shorter wavelength emission is clearly associated with ionising stars. Such SFR are much smaller than the ones obtained for our FIRBACK galaxies. A larger sample of galaxies has been observed in Hα by James et al. (2004) to derive SFRs: they find an integrated, extinction-corrected SFR of 1-3 $M_\odot$/yr for the most active spirals, the types Sbc or Scs. Our Hα rates are difficult to compare directly because of the uncertain extinction correction: but either applying the average correction of 10 discussed earlier, or using individually corrected Hα rates from the similar, southern sample discussed by Patris et al. (2003), or using the FIR-SFR which correlate well with the corrected Hα rates as shown by Patris et al. (2003), we have here SFRs on average 3 to 4 times greater than the most active galaxies of James et al. (2004), the Scs. Furthermore, our morphologies are generally of earlier types than Sc (with often a central, bulky component), types for which the SFRs measured in James et al. (2004) are comparatively smaller. By comparison, M101, a nearly face-on spiral whose SED, shown in Fig. 5, is comparable to the quiescent one of Dale et al. (2001), has a FIR-SFR of 5.7$M_\odot$/yr (for an IR luminosity of 4.4 10^{10} $L_\odot$, derived from IRAS data), similar to the ones of our FIRBACK galaxies. But its classification is SBc, again of later type than most of our galaxies, and furthermore with a bar, usually believed to be linked with larger SFRs than in non-barred galaxies.

The IR luminosity function derived from the Bright Galaxy Sample by Soifer et al. (1987) shows that the break occurs at a luminosity of 1.7 10^{10} $L_\odot$. Following Stickel et al. (2000), this value, calculated over the 40-220 μm range, converts into roughly 3 10^{10} $L_\odot$ for the 3-1000 μm range discussed in this paper: the majority of our objects are therefore typical L^* objects, or slightly above. Noticable exceptions are FN1-0, 5, 15, 39, 40 or 43, which are much more luminous. When comparing with the local K-band luminosity function of Cole et al. (2001), most of our galaxies are only slightly brighter than the M_K^* magnitude of -24.2 (with H_o = 70 km/s/Mpc as used here) (the exceptions being FN1-1, 12, 23 and 40 which are significantly fainter), so that they are not extremely massive galaxies. The ratio of total far-IR luminosity to the K-band luminosity, which can be interpreted as an indicator of the ratio of present to past star

formation, is rather homogeneous for the sample, with values between 20 and 40, showing that the present SF dominates the energy output. A few objects are more active than average (FN1-1, 23, 39, 40, 41), but are also less massive (with the exception of FN1-39). The outstanding object is FN1-40 (which is also the most distant object of the sample), appearing as a sub-L* object from its K magnitude, but heavily forming stars, and with a far-IR luminosity bringing it into the ULIRG class: this galaxy is therefore clearly of a different nature than the average bright FIRBACK galaxy and is probably a dwarf galaxy experiencing one of its first starbursts; see also Chapman et al. 2002. We conclude that, while a few are standard, disk-dominated spirals (like for instance FN1-02), many of these objects seem to have a larger SFR than standard spirals, with a concentration towards the central regions which could indicate the final phase of a former merger event.

As far as the CIB is concerned, it is clear that the bright FIRBACK galaxies have mid- to far-IR colors clearly different from the mean colors of the CIB. They do therefore not represent the bulk of the sources contributing to the mid- and far-IR CIB, contrary to earlier expectations (e.g. Puget et al. (1999), or Devriendt & Guiderdoni (2000)). The main contributors to the IR-CIB are thus expected to be more distant, more active galaxies, which will possibly be detected within the fainter part of the FIRBACK survey.

This local cold population however has to be taken into account in the modelling of the evolution of IR galaxies. In practice, it strongly affects the redshift distribution of 170 μm sources predicted by the models. When models consider only starburst galaxies, they lead to a redshift distribution for the 4σ FIRBACK galaxies that is clearly biased towards redshifts much higher than those observed here (e.g. Devriendt & Guiderdoni (2000)). Our present results show that the number counts are dominated by local cold galaxies for 170 μm fluxes greater than about 240 mJy. This is now well taken into account in the Lagache et al. (2003) models, which predict an equal contribution of starburst- and of "cold" galaxies at 250 mJy. This is also in agreement with the statistical results derived from sub-mm observations by Sajina et al. (2003).

The detailed analysis of the FIRBACK population obviously needs to be completed by the identification of the fainter, more distant counterparts. This will require many complementary observations in various wavelength ranges, from the ground with larger facilities, and from space. But, as the confusion limit of the FIRBACK survey clearly restricts the detection to relatively nearby objects, results from *SPITZER* should notably improve the situation. Indeed, the first number counts at 24 μm in the N1 field by e.g. Chary et al. (2004) suggest a strong contribution from luminous IR galaxies in the redshift range between 0.5 and 2.5. Inclusion of the shorter wavelengths data from IRAC will also allow to better determine the contribution of the quiet, diffuse component to the overall energy output in the nearby objects. A detailed study of those

galaxies (which is underway; e.g. Sajina et al., in preparation) is therefore likely to bring new light on the evolution of IR galaxies and their relation to the higher-z sources found in sub-mm observations.

Acknowledgements: We thank Jean-Loup Puget for enlightening discussions and support during this work; P. Chanial for providing the M101 data in advance of publication; S.C. Chapman for providing us two redshifts (FN1-5, 39) and D. Dale for providing his SED models. S.M. acknowledges support from the ESA External Fellowship program, and advice from Marc Postman for the morphological classification of the galaxies.
This paper is based on observations made with ISO, an ESA project with instruments funded by ESA Member States and with the participation of ISAS and NASA.
This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by NASA and NSF. It also used data from the Lyon Extragalactic Database (LEDA).

References

Babbedge T.S.R., Rowan-Robinson M., Gonzalez-Solares E. et al., 2004, MNRAS 354, 961
Basilakos S., Georgantopoulos I., Pérez-Fournon I. et al., 2002, MNRAS 331, 417
Becker R.H., White R.L. & Helfand D.J., 1995, ApJ 450, 559
Bertin E., Dennefeld M. & Moshir M., 1997, A&A 323, 685
Chanial P. 2003, Ph. D. Thesis, University of Paris-7
Chapman S. C., Smail I., Ivison R.J. et al., 2002, ApJ 573, 66
Chary R., Casertano S., Dickinson M.E. et al., 2004, astro-ph 0406386
Ciliegi P., McMahon R.G., Miley G. et al., 1999, MNRAS 302, 222
Cole S., Norberg P., Baugh C.M¿ et al., 2001, MNRAS 326, 255
Condon J.J., Cotton W.D., Greisen E.W. et al., 1998, AJ 115, 1693
Condon J. J., 1992, Ann. Rev. Astron. Astrophys. 30, 575
Condon J. J. & Broderick J. J., 1991, AJ 102, 1663
Dale D., Helou G., Contursi A. et al. 2001, ApJ 549, 215
Dale D., & Helou G., 2002, ApJ 576, 159
Devriendt J., Guiderdoni B. & Sadat R. 1999, A&A 350, 381
Devriendt J. & Guiderdoni B. 2000, A&A 363, 851
Dole H., Gispert R., Lagache G. et al. (FIRBACK III), 2001, A&A 372, 364
Elbaz D., Cesarsky C., Chanial P. et al. 2002, A&A 384, 848
Flores H., Hammer F., Désert F.-X. et al. 1999, A&A 343, 389
Gonzalez-Solares E.A., Perez-Fournon I., Rowan-Robinson M. et al. 2005, MNRAS in press, astro-ph 0402406
Hauser M.G., Dwek E., 2001, ARAA, 37, 249
Helou G., Soifer B. T., & Rowan-Robinson M. 1985, ApJ 298 L7
Helou G., Roussel H., Appleton P. et al., 2004, ApJ Supp. 154, 253

Heraudeau Ph., Oliver S., del Burgo C. et al., 2004, MNRAS 354, 924

Hinz J. L., Rieke G. H., Gordon K. D. et al., 2004, ApJ Supp. 154, 259

James P. A., Shane N. S., Beckman J. E. et al., 2004, A&A 414, 23

Kawara K., Sato Y., Matsuhara H. et al. 1998, A&A 336, L9

Kennicutt R.C. 1992, ApJ 388, 310

Kessler M. F., Steinz J. A., Anderegg M. E. et al. 1996, A&A 315, 27

Kleinmann S. G., Lysaght M. G., Pughe W. L. et al. 1994, Ap&SS 217, 11

Lagache G., Dole H. & Puget J.L., 2003, MNRAS 338, 555

Lagache G., Dole H. & Puget J.L., et al., 2004, APJSS, in press

Lilly S.J., Carollo C.M. & Stockton A.N., 2003, ApJ 597, 730

Lonsdale C. J., Smith H. E., Rowan-Robinson M. et al. 2003, PASP 115,897

McMahon R. G., Walton N. A., Irwin M. J. et al. 2001, New A.R., 45, 97

Mann R.G., Oliver S., Carballo R. et al. 2002, MNRAS 332, 549

Manners J.C., Serjeant S., Bottinelli S. et al. 2004, MNRAS 355, 97

Oliver S., Rowan-Robinson M., Alexander D.M. et al. 2000, MNRAS 316, 749

Patris J., Dennefeld D., Lagache G., Dole H., 2003, A&A 412, 349

Pettini M., Shapley A.E., Steidel C.C. et al., 2001, ApJ 554, 981

Postman M. et al., 2005, ApJ submitted

Puget J.-L., Abergel A., Bernard J.-P. et al., 1996, A&A 308, 5

Puget J.-L., Lagache G., Clements D.L¿ et al. 1999, A&A 345, 29

Renault C., Barrau A., Lagache G. and Puget J.-L., 2001, A&A 371, 771

Rowan-Robinson M., Lari C., Perez-Fournon I. et al., 2004, MNRAS 351, 1290

Sajina A., Borys C., Chapman S. et al., 2003, MNRAS 343, 1365

Scott D., Lagache G., Borys C. et al., 2000, A&A 357, L5

Soifer B. T., Sanders D. B., Madore B. F. et al., 1987, ApJ 320, 238

Soifer B. T., Boehmer L., Neugebauer G. & Sanders D. B., 1989, AJ 98, 766

Stickel M., Lemke D., Klaas U. et al., 2000, A&A 359, 865

Stickel M., Lemke D., Klaas U. et al., 2004, A&A 422, 39

Strateva I., Ivezic Z., Knapp G.R. et al., 2001, AJ 122, 1861

Surace J. A., Shupe D. L., Fang F. et al., 2004, http://ssc.spitzer.caltech.edu/legacy/

Vaccari M., Lari C., Angeretti L. et al., 2005, MNRAS in press, astro-ph 0404315

Veilleux S., Kim D.C., Sanders D.B. et al., 1995, ApJ 98, 171

Veilleux S., 2002, IAU Coll. 184, R.F. Green, E.Y. Khachikian and D.B. Sanders, Edts, ASP Conf. Series 284, 111

Werner M. W., Roellig T. L., Low F. J. et al., 2004, ApJ Supp. 154, 1

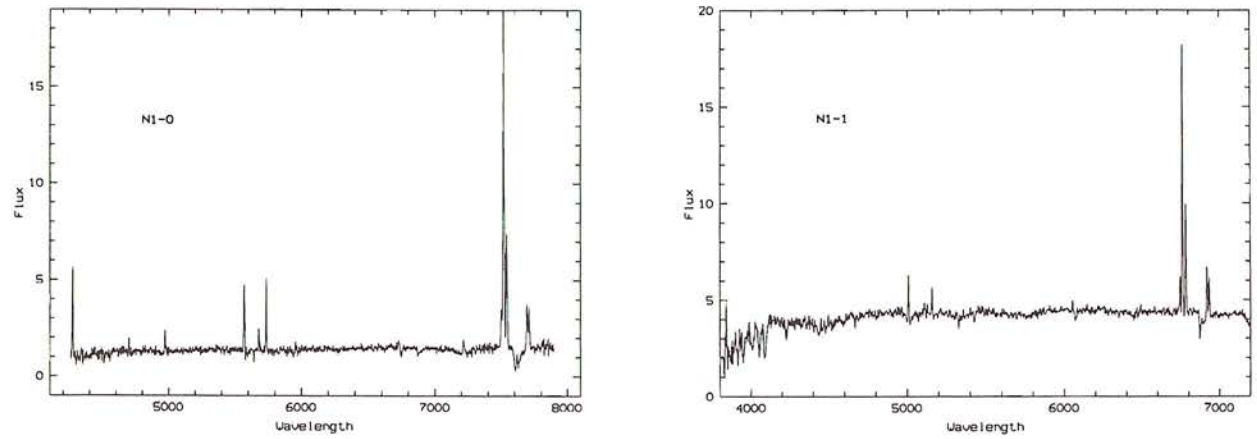

Fig. 9. Spectra of individual objects as examples (FN1-0 and FN1-1). Spectra of further objects are only available in electronic form.

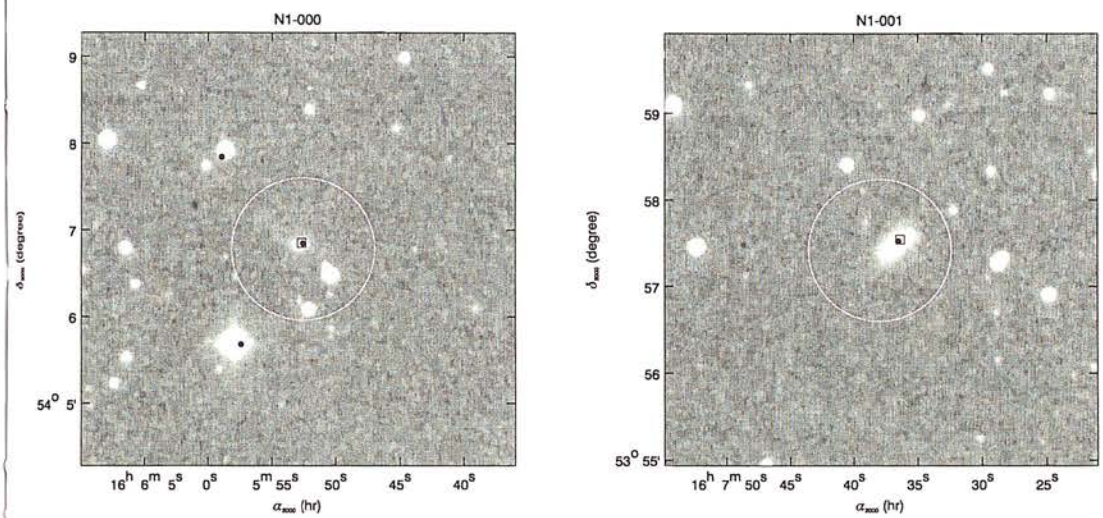

Fig. 10. Identification charts of FIRBACK N1 field sources (FN1-0 and FN1-1 as examples). Charts of further objects are only available in electronic form. The error circle for the 170μm ISO source is superposed on the DSS optical image. An ISO 15μm source is marked by a black dot and a radio source by a small open black square.